JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 21, No. 8 AUGUST, 1950

In This Issue

		Page
Conductivity of Liquid Selenium—200° 500°C	HERBERT W. HENKELS	725
Specific Heats of MgO, TiO_2, and ZrO_2 at High Temperatures	JAMES S. ARTHUR	732
Effect of a Video Filter on the Detection of Pulsed Signals in Noise	DAVID MIDDLETON	734
Negative Wire Corona at High Temperature and Pressure	L. R. KOLLER AND H. A. FREMONT	741
Radiation from a Transverse Rectangular Slot in a Circular Cylinder	SAMUEL SILVER AND WILLIAM K. SAUNDERS	745
Thermal Conductivity of Soils	ANDREW GEMANT	750
Theory of Blind Navigation by Dynamical Measurements	J. J. GILVARRY, S. H. BROWNE, AND I. K. WILLIAMS	753
Diffraction Pattern in a Circular Aperture Measured in the Microwave Region	C. L. ANDREWS	761
Retrograde Motion of the Arc Cathode Spot	C. J. GALLAGHER	768
Transonic Potential Flow of a Compressible Fluid	W. R. SEARS	771
Analysis of Intensities Obtainable Using Pinhole and Slit Collimators in X-Ray Powder Diffraction Techniques	LEROY ALEXANDER	779
Average Electron Density Measurements by Low Angle X-Ray Scattering	H. M. BARTON, JR. AND R. BRILL	783
Removable Intermediate Lens for Extending the Magnification Range of an Electron Microscope	J. HILLIER	785
Wall-Quenching of Laminar Propane Flames as a Function of Pressure, Temperature, and Air-Fuel Ratio	RAYMOND FRIEDMAN AND W. C. JOHNSTON	791
Electromagnetic Field of the Conical Horn	MARVIN G. SCHORR AND FRED J. BECK, JR.	795
On the Diffraction of a Radar Wave by a Conducting Wedge	R. B. WATSON AND C. W. HORTON	802
Microscopic Observation of the Solidification of Small Metal Droplets	D. TURNBULL AND R. E. CECH	804
Mass Spectrometric Study of Solids. I. Preliminary Study of Sublimation Characteristics of Oxide Cathode Materials	R. H. PLUMLEE AND L. P. SMITH	811
Static Electrification of Dust Particles on Dispersion into a Cloud	W. B. KUNKEL	820
Charge Distribution in Coarse Aerosols as a Function of Time	W. B. KUNKEL	833
Double Crystal and Slit Methods in Small Angle X-Ray Scattering	H. N. RITLAND, P. KAESBERG, AND W. W. BEEMAN	838
Letters to the Editor:		
Vacuum Tubes with Mutually Bombarding Oxide Cathodes	E. G. HOPKINS	841
Electron Optical Properties of Space-Charge Clouds	L. MARTON AND D. L. REVERDIN	842
Index to Advertisers		xxvi

BOARD OF EDITORS
ELMER HUTCHISSON, *Editor*

Associate Editors
HOWARD H. AIKEN
OTTO BEECK
JAMES HILLIER
W. JAMES LYONS
PHILIP MORRISON
ERIC REISSNER
RAYMOND J. SEEGER
FREDERICK SEITZ
CHARLES S. SMITH

Publication of the
American Institute of Physics

OFFICERS
GEORGE R. HARRISON, *Chairman*
GEORGE B. PEGRAM, *Treasurer*

GOVERNING BOARD
J. W. BEAMS
C. PAUL BONER
WALLACE R. BRODE
W. F. FAIR, JR.
J. B. FISK
G. P. HARNWELL
GEORGE R. HARRISON
W. V. HOUSTON
RUDOLPH KINGSLAKE
PAUL KIRKPATRICK
PAUL E. KLOPSTEG
F. W. LOOMIS
PHILIP M. MORSE
J. R. OPPENHEIMER
GEORGE B. PEGRAM
HAROLD K. SCHILLING
FRANCIS G. SLACK
JOHN C. STEINBERG

Consultant on Printing
MELVIN LOOS

ADMINISTRATIVE STAFF
HENRY A. BARTON, *Director*
WALLACE WATERFALL, *Executive Secretary*
THEODORE VORBURGER, *Assistant Treasurer and Comptroller*
RUTH F. BRYANS, *Publication Manager*
BETTY HOFSTADTER, *Assistant Secretary*
EILEEN B. NEUBERGER, *Circulation Manager*

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 21, Number 8 August, 1950

Conductivity of Liquid Selenium—200°–500°C*

HERBERT W. HENKELS
Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pennsylvania
(Received October 20, 1949)

The conductivity of liquid selenium has been measured in the temperature range 200°–500°C. The resistivity was expressed by $\log_{10}\rho = A + (B/T)$. Average values for different selenium lots and melts of A and B were -3.81 and 5850. The maximum deviations from the averages were 10 percent and 3.4 percent respectively. The resistivity was a function of temperature alone.

Various non-metal impurities Cl_2, I_2, P lowered the resistivities and produced different values of B in different temperature ranges. Mercury addition caused no change in either A or B although that metal greatly influences the resistance of solid hexagonal selenium. Melts doped with Cl_2 or Br_2, along with Hg exhibited behaviors different from those with single additions.

It was concluded that selenium is an ideal semiconductor in the range of measurement.

INTRODUCTION

STUDIES of the electrical properties of selenium are in progress at the University of Pennsylvania. Investigations on the subject have been summarized in Technical Report No. 9 entitled "Summary of Recent Work on the Conductivity of Selenium." It develops from a consideration of the data presented that a satisfactory explanation of conduction and the absence of measurable Hall effect[1] in solid selenium may require data on the conductivity of pure and doped, liquid and amorphous selenium. Existing data is scarce and unsatisfactory for reasons to be discussed later.

The study was considered profitable for several additional reasons. Satisfactory solid samples must at present be made according to an entirely empirical procedure. Studies of solid samples are greatly influenced by the crystalline condition of the sample. Therefore a detailed examination of the conditions of the selenium in all stages of sample preparation seemed highly desirable. Selenium is the simplest of the materials forming a stable amorphous phase. The data presented may prove useful in the further development of the theory of the liquid state. Commercial rectifiers are prepared from selenium doped with different halides. A continuous indication of the percentage impurity present in the molten selenium batch would be useful. The present report shows the possibilities of such an indication.

EXISTING DATA

Pelabon[2] reported a single measuring series with temperature variation from 390 to 690°C. The data was represented by:

$$\log\rho = 8.506 - 0.0095t(°C).$$

At 300° the resistance was noticed to decrease suddenly with a velocity that made measurements impossible. The origin of the selenium used in the experiments was not mentioned.

Borelius *et al.*[3] presented data on the resistance of solid and liquid selenium. Various selenium samples were sealed in glass tubes equipped with suitable platinum electrodes and the resistance was measured both as functions of periods of heating and of temperature. In the range measured, the specific resistance ρ followed the law, $\log\rho = A + (B/T)$, where T is the absolute temperature, A and B are constants. The results are given in Fig. 12.

* The following work has been done under contract NObs 34144 of the Bureau of Ships, United States Navy Department and previously reported as Technical Report No. 13 for that contract.

[1] H. W. Henkels, "Summary of Recent Work on Conductivity of Selenium," BuShips Contract NObs-34144, Technical Report No. 9, March 5, 1948.

[2] H. Pélabon, "Resistivity of selenium," Comptes Rendus **173**, 295 (1921).

[3] Borelius, Pihlstrand, Anderson, and Gullberg, "Resistivity of selenium," Arkiv. Mat. Astron. Fysik **30A**, No. 14 (1944).

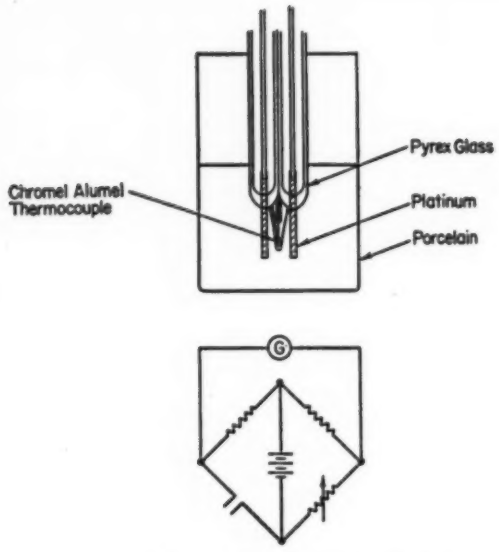


FIG. 1. Conductivity cell (liquid 200–500°C).

The values of A and B were observed to vary with the sample, period of heating and with depth of the electrodes in the melt. The heating periods were quite long (about 1800 hours). On the basis of the variation of resistance with depth of electrodes and differences observed in the conductivity of the cooling products of different strata of the melt, two liquid phases were postulated, a heavier one with low value of B and a lighter one with higher value of B. The properties of liquid selenium were adjudged similar to those of liquid sulfur.

NEW DATA

Source of Selenium

Selenium for the experiment was obtained from the Canadian Copper Refiners, Ltd. The analysis of the C.P. shot, performed by the laboratory of the company, was as follows:

Cu	Fe	Te	N.V.R.
0.1	0.8	0.2	5

parts per million. The selenium was purified in the following steps:†

1. Distillation at atmospheric pressure;
2. Oxidation in vapor state to SeO_2;
3. Resublimation of the SeO_2, followed by reduction in vapor state with NH_3 to elemental selenium;
4. Distillation of selenium in quartz apparatus, followed by condensation and shotting in distilled water.

TABLE I.

Sample	Curve	Temp.	A	B
lot 319 melt 1	$A, B, C,$	250–500°C	—	5650
lot 319 melt 2	D, E, F, G	250–500°C	—	5860
lot 319 melt 2	H, I	215–500°C	−3.99	5900–6010
lot 319 melt 3	N	250–500°C	−3.87	5930
lot 319 melt 4	W, X, Y	250–500°C	—	5930
lot 504–2 melt 1	AG	200–500°C	−3.50	5630
lot 504–2 melt 1	AH	200–500°C	−3.88	5882

† Information from Dr. C. W. Clark, Director of Research and Development, Canadian Copper Refiners, Limited.

Experimental Arrangements

Resistance was measured with a conductivity cell in an ordinary d.c. bridge circuit.

The conductivity cell is sketched in Fig. 1. Forty mil platinum wires were sealed in Pyrex tubing with 15 mm exposed. The glass tubes were then fastened together so that the electrodes were spaced 9 mm apart. A third glass tube with the end drawn to a small diameter and sealed, was fastened to the other tubes and served as the well for a thermocouple. This arrangement was placed in the center of a small porcelain beaker containing the sample which in turn was situated in the center of an electrically heated oven.

Decade resistance boxes were used as the arms of the bridge. These were calibrated and were found to have an error less than 1 percent. A Weston galvanometer with sensitivity of 0.25×10^{-6} amp./mm was used as null indicator.

For temperature measurements a chromel alumel thermocouple was introduced into the thermometer well. One wire was threaded into a piece of fine tubing to prevent shorting after the insulation had melted. The thermocouple hot junction was in contact with the thin drawn walls of the well. The e.m.f. was measured with a type K potentiometer whose smallest scale division was 5 microvolts. A Rubicon galvanometer with sensitivity

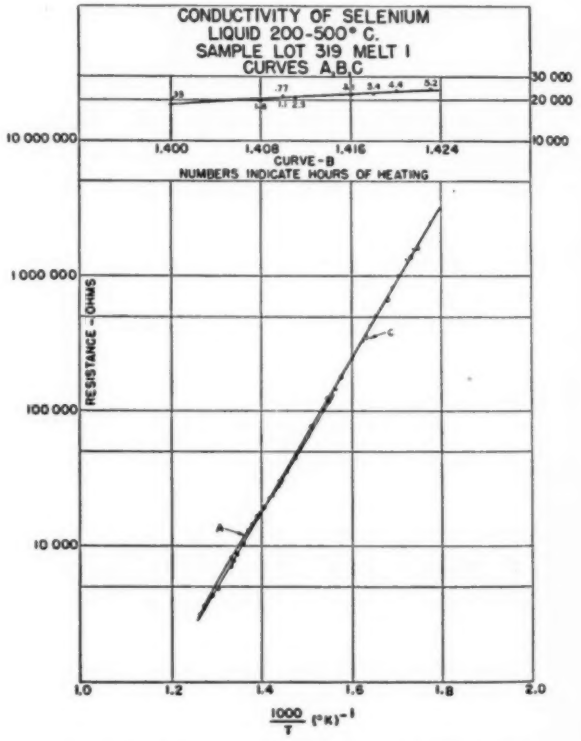


FIG. 2. Resistance of pure selenium, lot 319, melt 1.

of 6×10^{-10} amp./mm was used for null indicator. The standard e.m.f. was furnished by a new Eppley cell recently calibrated. Cold junction temperature was measured with a small mercury thermometer.

The conductivity cell constants were determined by calibration against an 0.01N KCl solution. The depth of the electrodes in the solution did not affect the value. The cell constants were the order of 1 ohm cm/ohm.

Consideration of Contact Resistance

Measurements on a cell with variable distance between the electrodes showed that contact resistance is of no importance.[3] This result agrees with the observations of Schweickert and Ioffe on the connection between

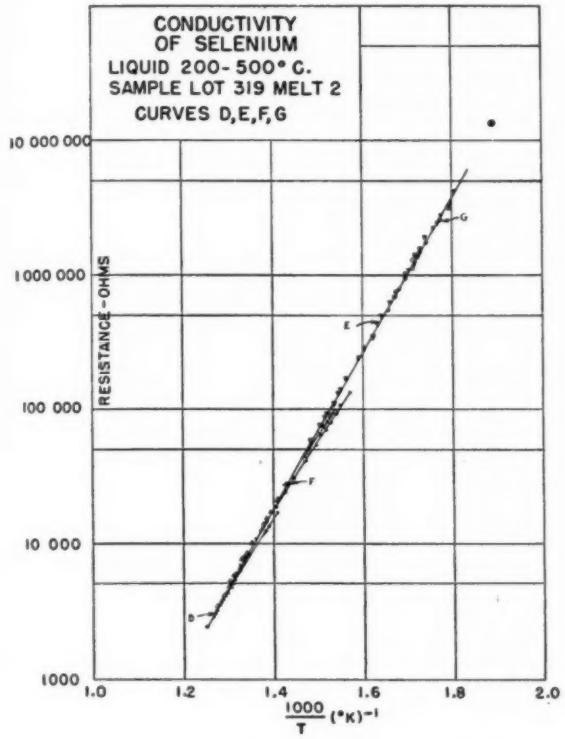


FIG. 3. Resistance of pure selenium, lot 319, melt 2.

contact resistance of metals against selenium and the work function of the metals.[4,5]

Procedure

The cell was carefully washed, rinsed in absolute methyl alcohol and dried in an oven. Selenium shot was introduced and melted. The heating procedure was varied from run to run in order to establish the dependence, if any, of the resistance on previous treatment. The thermocouple e.m.f. was measured simultaneously with the resistance. Care was taken to avoid contaminating the selenium with dust. Any contamination

[4] Schottky, Zeits. f. Physik **118**, 539 (1941).
[5] A. V. Ioffe and A. F. Ioffe, "Contact of semiconductors with metals," Bull. Acad. Sci. Ser. Phys. **5**, 550 (1941).

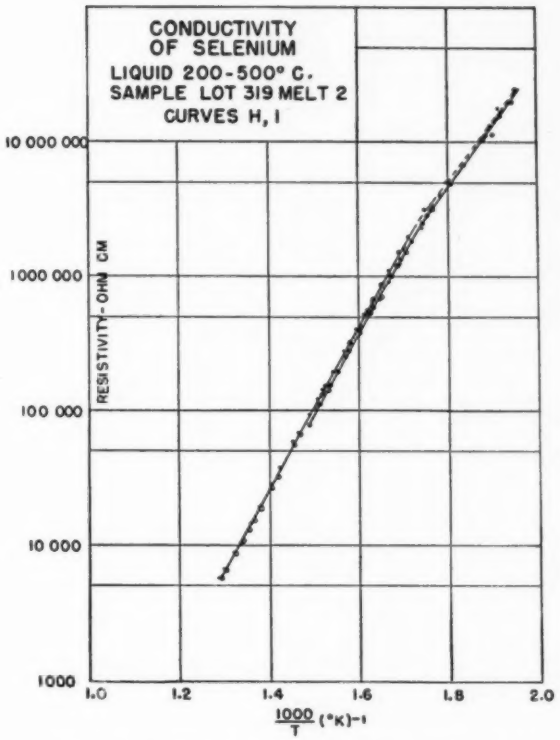


FIG. 4. Resistivity of pure selenium, lot 319, melt 2.

was evidenced by the presence of small crystallized spots in the amorphous selenium surface after cooling slowly to room temperature.

Accuracy of Measurements

The precision of measurement of resistance was limited by the spread of the null indication. At low resistances the error was less than 1 percent, at high values (10 megohms) perhaps as high as 5 percent. The error in reading temperatures was about ±1.00°C arising from difficulty in balancing the bridge and the potentiometers simultaneously. The cell calibration was accurate to ±5 percent.

Data

The data collected is plotted in curves in Figs. 2–12. The results are reviewed in the discussion.

DISCUSSION

Pure Selenium

Care was taken in assembling the data to attain equilibrium conditions. Time rates of change of temperature varied from run to run but averaged about two or three degrees per minute. This rate gave satisfactory curves throughout most of the range. Below 300° a slower rate was desirable. Curves were retraced on reversing the sign of the temperature change. At low temperatures hysteresis loops occurred in a few cases (see curves *N*, *O*, *Q*, *V*, *AE*).

TABLE II.

Sample	Curve	Temp.	A	B
lot 319 melt 2	H, I	215–500	−3.99	5900–6010
lot 319 melt 2 1% I_2	J	215–500	−1.79	3980
I_2 J	K	215–500	−1.75	4020
I_2 K	L	215–500	−1.96	4240
I_2 L	M	215–500	−2.12	4420

TABLE III.

Sample	Curve	Temp.	A	B
lot 319 melt 3	N	250–500	−3.87	5930
lot 319 melt 3 with 0.1% mercury	O	250–500	−3.87	5930
lot 319 melt 3 with 0.1% mercury	Q	250–500	−4.00	6040

Curves $A, B, C, D, E, F, G, I, N, V, W, X, Y, AG, AH$ give data on pure selenium. Temperature dependence of resistance followed the law

$$\log_{10}\rho = A + B/T$$

with values of A and B given in Table I. An average of the values of B gives 5850 with maximum deviation 3.4 percent, of A gives −3.81 with 10 percent deviation.

Curves $H, I, O, Q, V, W, X, Y, AD, AE$, perhaps N, AG and AH show a slight change in temperature coefficient about 300°C. The kink is most easily noticeable in experiments with selenium that has been exposed to air a considerable length of time. This change has been attributed to the presence of SeO_2.

Curve B shows the effect of a short period of heating at about 400°C on the resistance of pure selenium. No change in resistance is apparent. A study of temperature coefficients, B, in Table I, does reveal a slight change in slope with prolonged heating but this is almost within the experimental error of the procedure. The curves AD, AE, AF of Fig. 10 of results on pure melts 5 and 6 give rather low temperature coefficients ($B=5200$ in curve AD). Incidentally, the kink at about 300° is well shown in AE. It was evident from the appearance of crystallized portions in the melt on cooling in the oven that these samples were contaminated. Sample AF completely crystallized on cooling. After the experiment was completed the entire apparatus (oven, hood, etc.) was cleaned thoroughly. Lot 504–2 melt 1 then repeated the values of slope obtained in former samples.

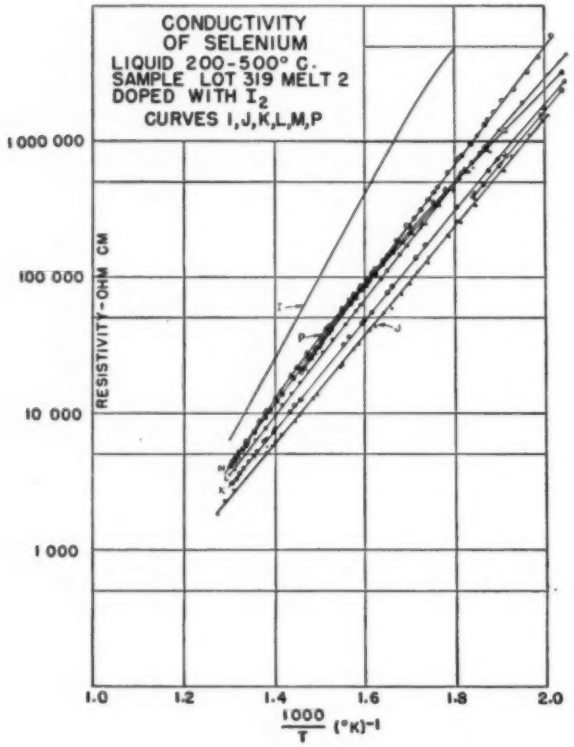


FIG. 5. Resistivity of I_2 doped selenium, lot 319, melt 2.

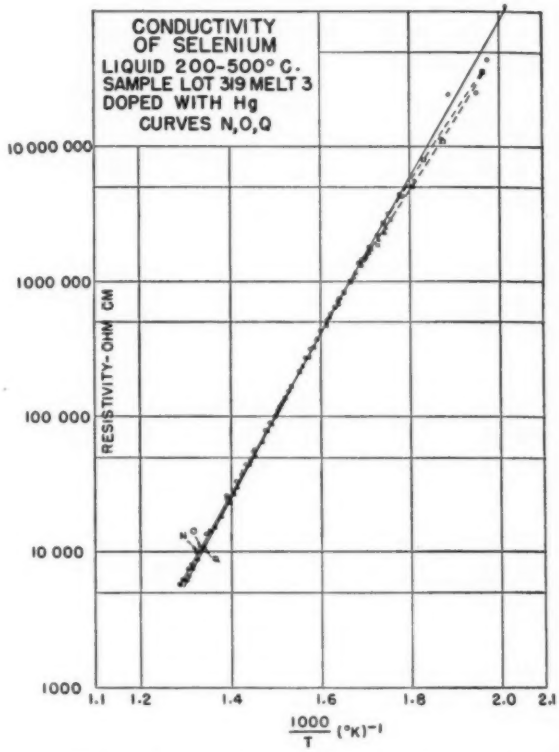


FIG. 6. Resistivity of Hg doped selenium, lot 319, melt 2.

The results of H Pélabon are plotted in Fig. 12. At 300° the resistance decreased suddenly with a velocity such that measurements were not possible. This observation was not checked by our data. At temperatures below 220° all samples crystallized (except those containing phosphorous) with a velocity depending on the amount and type of impurity present.

The data of Borelius is plotted in the same figure. Temperature coefficients varied considerably among samples of different purity (as would be expected) but also changed with different measurements on the same

sample and with depth of the conductivity electrodes in the melt. An equilibrium of two liquid phases of different densities and conductivities was postulated. In our data all pure samples gave the same results. No variation of resistance with depth was noticed. A. N. Campbell and S. Epstein[6] found no indication of shifting internal equilibrium in density measurements over the range 20–277°C. Dobinski and J. Wesolowski[7] found the density exhibited no singularities between 217° and 300°C. The viscosity temperature curve of selenium is given by Dobinski and Wesolowski[8] that of sulfur by

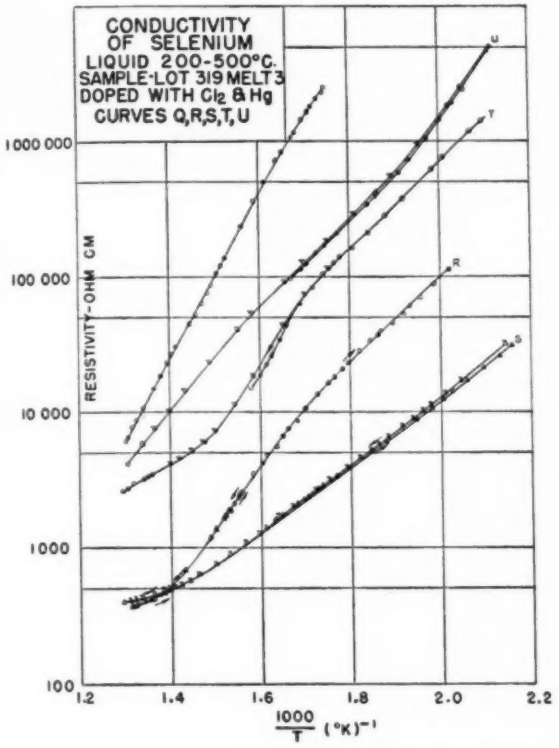


FIG. 7. Resistivity of selenium doped with Hg and Cl_2, lot 319, melt 3.

Bacon and Fanelli.[9] Selenium shows no tendency to form a ring structure at low temperatures as evidenced by the shape of the viscosity curve.

Doped Selenium

The melt of curve *I* was cooled to about 250°C and 1 percent by weight of iodine Se_2I_2(?) was introduced. After a mixing period of one minute, curve *J* was traced. Curve *K* was obtained after a $3\frac{1}{2}$-hour period of heating above 400°C, curve *L* after additional 5 hours, curve *M*

[6] A. N. Campbell and S. Epstein, "Density of selenium," J. Am. Chem. Soc. **64**, 2679–80 (1942).
[7] S. Dobinski and J. Wesolowski, "Density of selenium," Bull. Acad. Polon. **8–9A**, 446–50 (1936).
[8] S. Dobinski and J. Wesolowski, "Viscosity of Liquid Selenium," Bull. Acad. Polon. **49(A)**, 7–14 (1937).
[9] Bacon and Fanelli, "Viscosity of sulfur," J. Am. Chem. Soc. **65**, 639–48 (1943).

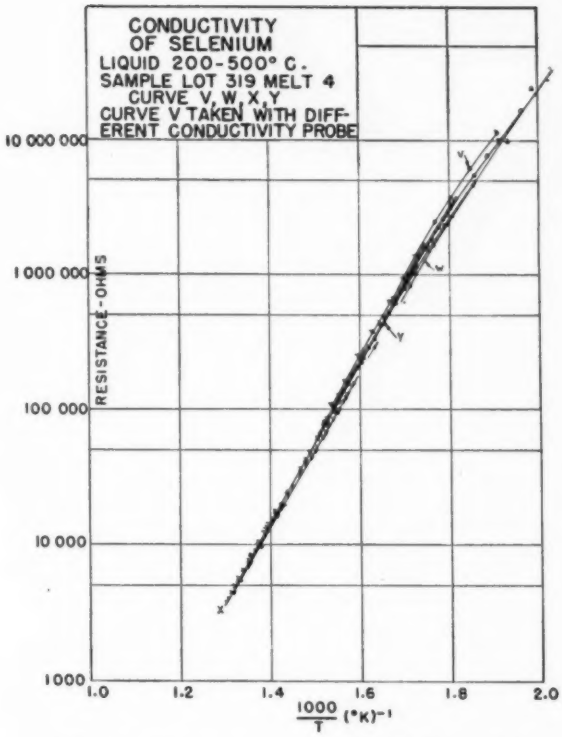


FIG. 8. Resistance of pure selenium, lot 319, melt 4.

after an additional $1\frac{3}{4}$-hours heating. Table II gives the changes in *A* and *B*. The temperature coefficients are observed to have increased slowly as I_2 was expelled.

0.1 percent mercury as HgSe was added to pure selenium of curve *N* and curve *O* was traced. No change in *A* or *B* was noted. The melt was then heated for 1 hour above 400°C and curve *Q* obtained. Again as indicated in Table III, no definite change in either *A* or *B* was observed.

Melt *M* was then doped with 0.1 percent mercury and immediately retested. Whereas no kinks were evident in curve *M* two appeared in curve *P* (see Table IV). Since conditions were the same in experiments *M* and *P* the changes in temperature coefficients at various temperatures must be attributed to the activity of various compounds of mercury and iodine. Mercury by itself has no effect on temperature coefficients as shown in curves *O* and *Q*. The possible significance of this fact will be discussed later.

The effect on an addition of 1 percent Cl_2 as Se_2Cl_2 to the melt of experiment *Q* is shown in curve *R*. The selenium solidified during addition of Se_2Cl_2 so that run *R* may have been conducted on an inhomogeneous sample. After the melt was heated to 500°C, curve *S* was obtained on cooling. Two temperature coefficients were measured for curve *S*, from 190 to 395°C, 2460, from 395–500°C, 1460. The effect of periods of heating on this melt are shown in Curves *T* and *U*. Different coefficients are noticed in various ranges. Compounds of mercury and chlorine could be active. To check the effect of

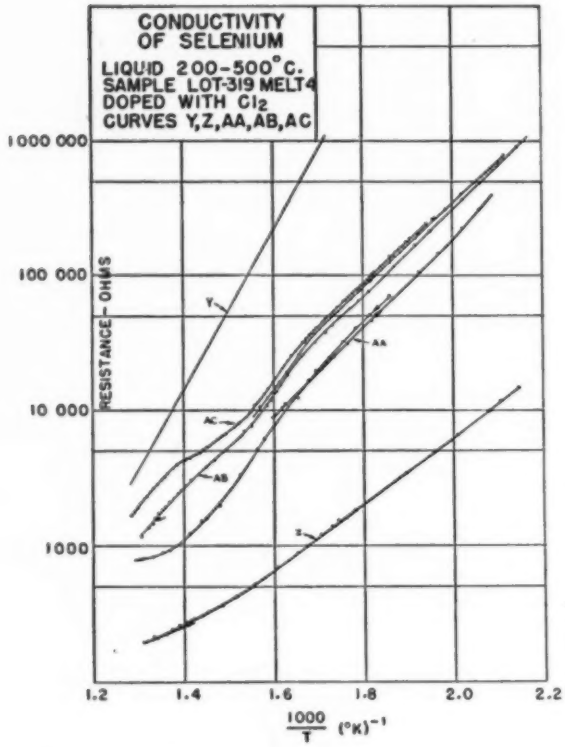


FIG. 9. Resistance of Cl_2 doped selenium, lot 319, melt 4.

chlorine, experiments Y, Z, AA, AB, AC were conducted on melts of pure selenium with Cl_2 added as Se_2Cl_2. This time the curve for low slope was obtained immediately. Temperature coefficients were almost identical in curves S and Z. (See Table V.) (Neither of these melts crystallized on cooling in the oven. This observation is explained in the following manner. Small amounts of impurities act as crystallization centers whereas excessive amounts inhibit the growth of crystals breaking up the continuity of crystallites as in the analogous case of oxygen in molten iron.) Curves AA, AB, AC have the same general appearance. In the last two cases a change in temperature coefficient seems to occur about 327°, one at 377° and another at 447°C. Interpretation of the curves is difficult. It is noteworthy that in all cases the resistances of the melts increased steadily with expulsion of chlorine by heating. Various compounds of platinum and chlorine and possibly oxygen influence the temperature coefficient. Experiments were conducted with melts containing 1 percent P. Equilibrium conditions were difficult to attain. A distinct change in temperature coefficient was noted at about 358°C. Below that temperature the value of B was 2170, above about 4200. This last figure is not very reliable. It is included to show the direction of the bend and the fact that the original value of B was never equaled or exceeded with impurities in the melt. (Melts with phosphorous would not crystallize under any condition. An explanation lies in the netting of molecular

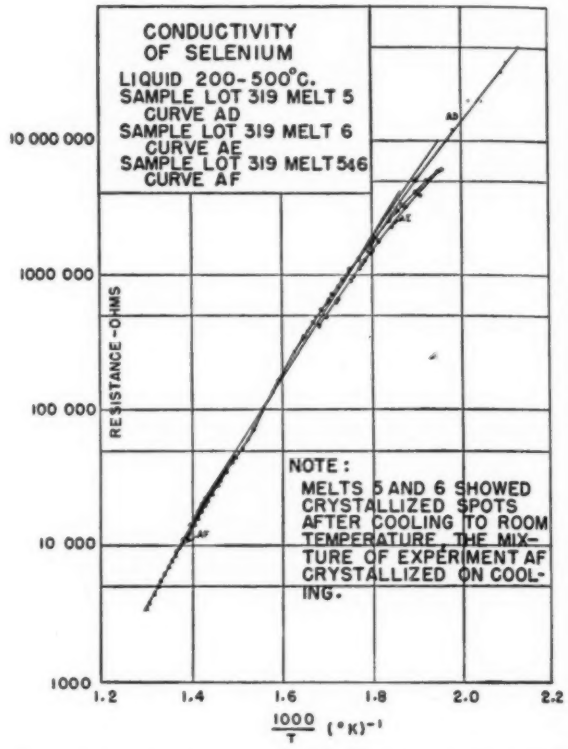


FIG. 10. Resistance of pure selenium, lot 319, melts 5 and 6.

TABLE IV.

Sample	Curve	Temp.	A	B
lot 319 melt 3 with I_2	M	215–500	−2.12	4420
lot 319 melt 3 with I_2 and mercury	P	215–275	−0.41	3400
		275–388	−1.05	3760
		388–500	−2.42	4650

TABLE V.

Sample	Curve	Temp.	A	B
lot 319 melt 3 with 0.1% mercury	Q	250–500	−4.00	6040
above with 1% Cl_2	S	190–395	−0.84	2460
		395–500	0.54	1460
lot 319 melt 4	W, X, Y	250–500	—	5930
above with 1% Cl_2	Z	190–395	—	2430
		395–500	—	1460

TABLE VI.

Sample	Temperature	A	B
Pure Selenium	200–500	−3.81	5850
Above with 1% I_2	215–500	−1.79	3980
Above with 1% Cl_2	190–395	−0.84	2430
	395–500	0.54	1460
Above with 1% P	225–358	—	2170
	358–500	—	ave 4240 (?)
Above with 0.1% mercury	250–500	−3.87	5930

threads caused by the trivalence of the element as discussed in Technical Report No. 9.)[1]

Data on different impurities in selenium is collected in Table VI.

The data indicates that selenium when pure, behaves as an ideal semiconductor in the range of temperatures measured. Non-metal impurities decrease the activation energy of the element in an amount depending on the type of impurity. Note the wide difference between the effect of iodine and chlorine, and to a smaller extent on the amount of impurity (experiments H, I, J, K, L, M). The impurity mercury was chosen as one usually considered to have a radical effect on the electrical properties of selenium. Experiments N, O, Q show that mercury was ineffective in changing the activation energies of the pure melt but was active in combining with the non-metals already in the selenium (curve P). Perhaps the discrepancy between the results of Penin and Astakhov[10] and rectifier manufacturers mentioned in Technical Report No. 9 can be resolved by assuming the experiments of the former were carried out on pure selenium whereas selenium doped with halogens was used commercially.

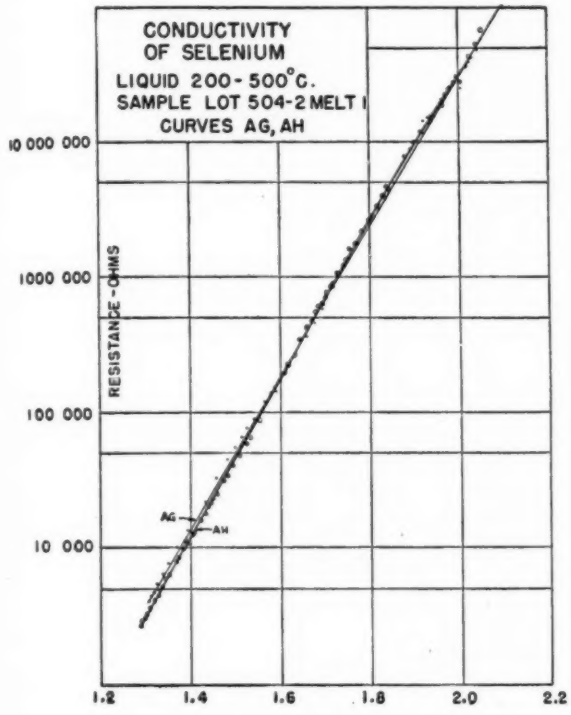


FIG. 11. Resistance of pure selenium, lot 504–2, melt 1.

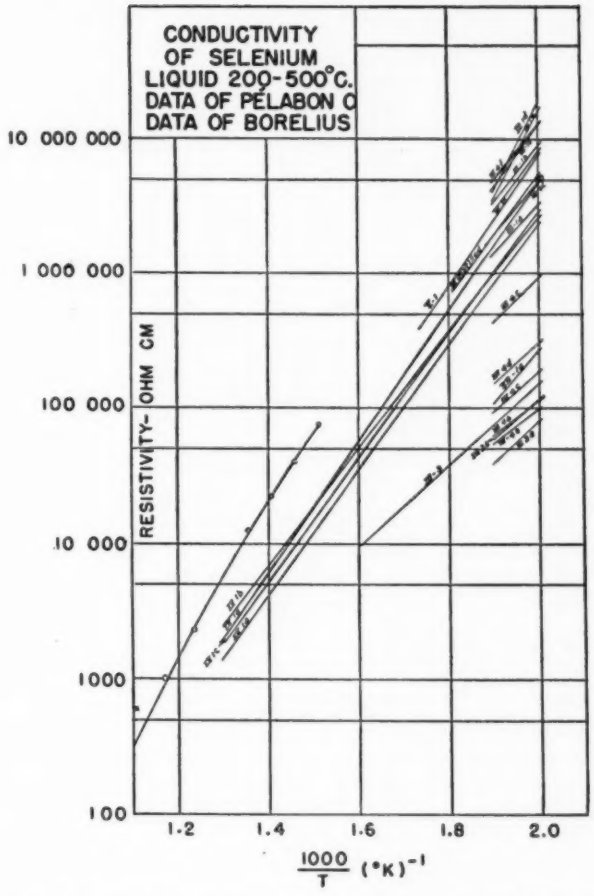


FIG. 12. Resistivity of selenium (data of Pélabon and Borelius).

Melts with metals and non-metals behave in a manner difficult to explain. It is however reasonable to assume that different compounds are stable in different temperature regions and that these compounds should influence the activation energies. There may perhaps be a certain amount of ionic conductivity. However the presence of a large thermoelectric power in liquid selenium and the value of activation energy indicates electronic conduction for the pure semiconductor.

[10] N. A. Penin and K. V. Astakhov, "Effect of admixtures to selenium on the electrical properties of selenium rectifiers," J. Tech. Physics (USSR) **16**, 199–206 (1946).

The Specific Heats of MgO, TiO_2, and ZrO_2 at High Temperatures*

JAMES S. ARTHUR
Department of Physics, The University of Pittsburgh, Pittsburgh, Pennsylvania
(Received June 23, 1949)

Mean specific heats were determined from 20°C to 800°C for MgO, TiO_2, and ZrO_2. True specific heats were then calculated assuming that $C_T = A + BT + CT^2$. At 350°C the true specific heats were calculated to be MgO 0.276; TiO_2 0.210; ZrO_2 0.140.

SOME data on the specific heats of oxides in the high temperature range, 20° to 1500°C, are available.[1,2] One of the methods of measuring specific heats employs an adiabatic calorimeter of the type described by Harry Hill and R. M. Bell.[3]

The calorimeter used in obtaining the data listed in this article was patterned after the one used by Hill and Bell with two main changes. An electrolytic, bath type heater was used instead of an electric coil heater in the jacket. The purpose of the heater is to enable the experimenter to match the temperature of the jacket surrounding the calorimeter cup with that of the cup and sample. The electrolyte used was tap water. Advantage is gained in getting more uniform heating and a much shorter reaction time. In addition, the plywood box enclosing the calorimeter cup was lined with aluminum foil to further reduce heat losses.

TABLE I. Mean specific heat of MgO.

Mean specific heat	t_2(°C)	t_1(°C)
Arthur		
0.2425	193.4	23.45
0.2466	197.0	24.25
0.2500	223.8	23.32
0.2519	259.2	25.75
0.2593	389.0	26.35
0.2590	398.7	26.18
0.2590	414.6	30.75
0.2657	483.8	30.08
0.2645	502.1	30.36
0.2733	582.8	31.75
0.2733	594.1	33.15
0.2687	595.2	35.81
0.2775	774.8	35.39
0.2806	800.7	37.00
0.2853	830.4	39.27
Wilkes		
0.2335	100	30.0
0.2470	300	30.0
0.2590	500	30.0
0.2690	700	30.0
0.2765	900	30.0

The calorimeter was calibrated by a small electric heating coil, as described by Hill and Bell.[3]

The heat capacity of the cup was found to increase with temperature in the range 20° to 40°C.

The oxide samples were obtained in powder form and packed into metal foil in the form of cylindrical pellets.[4] The dimensions of the pellets were slightly smaller than the dimensions of the cavity in the calorimeter cup. Aluminum foil was used for temperatures under 600°C and platinum foil when determinations were made at higher temperatures. A mold was used to make the pellets. It consisted of a cylinder composed of two close fitting pieces of hard wood with a brass insert, and a tamping piston made of dowel wood.

A Hoskins electric furnace, type FH204C was used to heat the samples. The temperature was regulated by a Brown electric pyrometer. All high temperature measurements were made with Chromel-alumel or iron-constanton thermocouples.

The methods used to calculate mean specific heat and true specific heat may be found in an article by Wilkes and Wood.[5] Values of mean specific heat are obtained directly from the data recorded and may be written as:

$$C_m = A + B/2(t_2 + t_1) + C/3(t_2^2 + t_2 t_1 + t_1^2), \qquad (1)$$

where C_m = mean specific heat, t_2 = furnace temperature, t_1 = final cup and sample temperature, and A, B, and C

TABLE II. Mean specific heats of TiO_2.

Mean specific heat	t_2(°C)	t_1(°C)
0.1720	123.2	24.10
0.1687	197.1	21.78
0.1748	200.1	24.30
0.1780	228.6	29.31
0.1838	289.2	27.02
0.1899	393.0	28.36
0.1930	400.6	29.56
0.1967	431.2	33.86
0.1959	455.8	31.56
0.1992	487.9	36.05
0.1929	491.8	31.50
0.1976	591.0	33.74
0.1965	593.4	34.12
0.2051	601.4	36.00
0.1967	602.5	32.50
0.2087	799.1	38.40

* From a thesis submitted in partial fulfillment of the requirements for the M.A. degree at Washington and Jefferson College.

[1] G. B. Wilkes, "The specific heats of MgO and Al_2O_3 at high temperatures," J. Am. Ceram. Soc. **15**, 72–7 (1932).

[2] Gronow and Schweite, "Die Spezifischen Warmen von CaO, Al_2O_3, $CaO \cdot Al_2O_3$, $3CaO \cdot Al_2O_3$, $2CaO \cdot SiO_2$, und $2CaO \cdot Al_2O_3 \cdot SiO_2$ von 20° bis 1500°C. Zeits. Anorg. allgem. Chem. **216**, 185–95 (1933).

[3] Harry Hill and R. M. Bell, "A versatile calorimeter for specific heat determinations," A.S.T.M. Bull. No. 151 March, 1948.

[4] Samples were obtained C. P. from Eimer and Amend.

[5] G. B. Wilkes and O. C. Wood, "the specific heat of thermal insulating materials, heating, piping and air conditioning," A. S. H. V. E. Journal Section, p. 370, June (1942).

TABLE III. Mean specific heats of ZrO_2.

Mean specific heat	t_2(°C)	t_1(°C)
0.1138	195.0	26.82
0.1163	197.4	23.20
0.1188	239.2	25.04
0.1222	278.7	26.62
0.1274	388.1	24.98
0.1275	398.1	29.17
0.1289	430.2	29.38
0.1308	488.9	31.17
0.1289	491.9	28.50
0.1332	579.8	32.56
0.1351	584.7	33.44
0.1332	606.0	32.64
0.1364	633.7	34.58
0.1363	795.7	40.26
0.1351	824.6	37.88
0.1437	943.4	42.87
0.1344	991.8	41.03

are constants. The true specific heat at any temperature is given by:

$$C_T = A + Bt + Ct^2, \qquad (2)$$

where C_T = true specific heat, t = temperature, and A, B, and C are the same constants as used in (1).

DISCUSSION OF RESULTS

The chief function of this article is to list the mean specific heats of MgO, TiO_2 and ZrO_2 in the range 20°–800°C and the true specific heats for the same oxides. The accuracy of the specific heats is judged to be ±2.0 percent.

MgO has a mean specific heat of 0.246 in the range 25°–200°C and 0.282 in the range 37°–800°C (see Table I). TiO_2 has a mean specific heat of 0.175 in the range 24°–200°C and 0.206 in the range 37°–800°C (see Table II).

ZrO_2 has a mean specific heat of 0.115 in the range 24°–200°C and 0.136 in the range 38°–800°C (see Table III).

Table I shows a comparison of the values of the mean specific heat of MgO obtained by the author with those obtained by G. B. Wilkes using an adiabatic bomb calorimeter. The author's values were consistently higher by approximately 0.3 percent for the same temperature range.

TABLE IV. Values of A, B, and C for MgO, TiO_2, and ZrO_2.

Material	$C_T = A + Bt + Ct^2$		
	A	B	C
MgO	0.2342	0.000124	−0.0000000170
TiO_2	0.1512	0.000245	−0.000000222
ZrO_2	0.0988	0.000173	−0.000000156

TABLE V. True specific heats of MgO, TiO_2, and ZrO_2.

Material	Temperature (°C)	True specific heat (Cal/g-deg. C.)
MgO	200	0.258
	350	0.276
	500	0.292
TiO_2	200	0.191
	350	0.210
	500	0.218
ZrO_2	200	0.127
	350	0.140
	500	0.146

Table IV is a list of the values of the constants A, B, and C which determine the equation of true specific heat in terms of temperature:

$$C_T = A + Bt + Ct^2.$$

Some values of C_T are listed in Table V. The values of true specific heat are accurate for temperature ranges approximately half that of the mean specific heats.

Errors resulting from variation in heat capacity of the cup with temperature are estimated at ±2.0 percent.

Radiation errors were minimized in the following ways: (1) A small refractory furnace was used to transfer the hot sample from the large furnace to the cup; (2) a baffle and guide arrangement was used to guide the sample into the cup and prevent direct heat radiation from the small furnace into the cup; (3) rapid transfers of samples were made, and mainly, (4) the use of the adiabatic heating jacket to keep the cup surroundings at the same temperature as the cup.

The Effect of a Video Filter on the Detection of Pulsed Signals in Noise*

DAVID MIDDLETON
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received November 30, 1949)

The effects of inserting a video filter of finite width on the observability of pulsed signals in random noise are examined. It is found that at match (when the pulse and IF filter are each other's conjugate Fourier transforms) no improvement is gained from such a video: the infinitely wide response yields the optimum results. This is strictly true when the second detector is a quadratic rectifier; (a slight improvement on narrowing the video is noted, however, for strong signals when a half-wave linear rectifier is used). Away from match a video filter does give noticeable improvement when the pulses are overlong, i.e., narrower spectrally than the IF ($\lambda>1$); more noise than signal is then removed by the narrower video filter. The greatest gain over the infinite video is observed for final filters slightly wider than the original pulse. On the other hand, for pulses that are too short ($\lambda<1$) the performance is worsened. The mean maximum signal level is decreased more rapidly than the interfering noise background. In all instances the familiar phenomenon of modulation suppression arises. Two different pulse shapes are considered: (1) Gaussian and (2) rectangular, but for wide videos pulse shape is not a critical factor. Curves showing the output signal-to-noise ratios as a function of the input ratio are included for conditions of match ($\lambda=1$) and mismatch ($\lambda\neq1$).

I. INTRODUCTION AND DISCUSSION

IN an earlier paper[1] the factors governing the reception of pulsed (radar) signals in random noise were discussed, and expressions for the output signal-to-noise ratio in terms of the input ratio were derived for three principal modes of perception: (1) observation of an oscilloscope screen, (2) listening to an harmonic of the pulse repetition frequency (PRF) or (3) noting the mean change in a meter's deflection in the presence and absence of a signal. Here, however, we are concerned with visual reception only, and it is the aim of this paper to examine the effects of inserting a finite video filter in the stage before observation (instead of the infinitely wide, uniform response assumed in the earlier analysis),[1] particularly when the pulse and IF filter are not matched, i.e., are not each other's conjugate Fourier transforms. Recent experiments by Fromm[2] indicate good qualitative agreement with the present theory. Some as yet unpublished,** theoretical discussion of the video filter and extensive experimental investigations, supporting our present conclusions were carried out during the war.[3]

The IF and succeeding stages of the receiver have a critical effect on reception. Coming into the IF is a fluctuating noise voltage and also a signal which is a known function of the time. After filtering and rectification the video amplitude cannot be represented as the simple sum of pure signal and "pure" (i.e., normal random) noise, as there are complicated modulation products generated between the two, especially following a linear detector. If the signal is repeated periodically there will be a certain interval during which the amplitude on the oscilloscope screen is on the average a maximum. Our definition, therefore, of what constitutes a signal in the visual case treated here is *the excess of the average of the amplitude, which we designate as* $s_v(0)_G$, *during this interval, over the mean amplitude* ($=n_v$, the r.m.s. video noise voltage) *in the absence of signal.* The concept of the average is needed because sometimes the noise interferes constructively and sometimes destructively with the signal. The above is called the "simple" criterion for visual detection[1] because, in practice, examination merely of the output ratio $s_v(0)_G/n_v$ neglects certain factors. For example, the effects of the (time) widths of the pulse *versus* the widths of the "blades of grass," or noise, are not considered, and in particular our definition does not take into account the action of pulse repetition: the more frequent the pulse, the more easily the eye integrates (in the limited time at its disposal) to detect the mean position of the pulse or pip with respect to the fluctuating background. To test the simple theory, therefore, we need a device, like the peak-reading meter,[2] which does not depend on PRF, but which gives $s_v(0)_G$ directly. The results are applicable to the radar problem within the limitations described above when the eye and brain constitute the final integrating elements. A table listing the various parameters—λ, Ω, Ω', involving the filter widths, pulse widths and durations, etc., that appear in the analysis—is included at the end of this section. Figure 1 illustrates the two types of pulse treated here and their respective spectra.

Figures 2–4 show the output video signal-to-noise ratio as a function of $\sigma_0^2\lambda$, where σ_0 is defined as the ratio of the r.m.s. maximum signal amplitude to the r.m.s. noise amplitude after passage through an IF filter which is matched to the incoming pulse [Eq. (2.13)]; for the Gaussian pulses and filter λ represents the ratio

* The research reported in this document was made possible through support extended Cruft Laboratory, Harvard University, jointly by the Navy Department (ONR) and the Signal Corps, U. S. Army, under Contract N5ori-76, T. O. 1.

[1] J. H. Van Vleck and D. Middleton, "A theoretical comparison of the visual, aural, and meter reception of pulsed signals in the presence of noise," J. App. Phys. 17, 940 (1946).

[2] W. E. Fromm, "An Experimental Investigation of the Reception of Pulsed Signals in the Presence of Noise," Thesis (Polytechnic Institute of Brooklyn, New York, May, 1948).

** *Note added in proof:* A similar analysis of the video filter problem is discussed in section 8.7 of reference 3 above. The results of both researches are in close and satisfactory agreement.

[3] J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals*, Book No. 24 (Massachusetts Institute of Technology) Radiation Laboratory Series, McGraw-Hill Book Company, Inc., New York, 1950.

of IF width to pulse width (frequency), cf. Eq. (2.15) and Table I. The factor λ is included to take into account the increased power necessary to give the same value of s_v/n_v at the output when the pulse and filter are no longer matched ($\lambda \neq 1$). Thus, Figs. 2–4 may be compared directly to show the well-known result[1] that at match a given value of the output ratio s_v/n_v (for the same arbitrary video filter width) is always obtained for the least energy per pulse or, equivalently, for the smallest average pulse power. Figures 2–4 also compare the Gaussian and rectangular pulses of initially the same energy [condition (2.23)]. At match for the Gaussian pulses ($\lambda=1$) the rectangular pulse is slightly off match, since our Gaussian IF is no longer the Fourier transform of the rectangular pulse. For this reason $(s_v/n_v)_{\text{rect.}\lambda=1}$ lies somewhat below the values of $(s_v/n_v)_{\text{Gauss.}\lambda=1}$ for the same video filters (same Ω).

Insertion of a finite video filter *at match* gives no improvement in the Gaussian case, provided the signal is weak compared to the noise, i.e., provided that the linear detector is functioning essentially like a quadratic one.[4] When the signals are strong ($\sigma_0^2 > 10$ db), however, a slight improvement is observed (Fig. 3) for video widths of the order of the original pulse width, i.e., $\Omega \sim 1$, 2. This occurs because a video filter helps to decrease the effects of the low frequency ($n \times n$) noise components arising from the higher-order modulation products generated in the linear detector, products which are absent in the output of the square-law rectifier. Insertion of a video filter away from match, however, will give significant improvement when the pulse is sufficiently long, a fact verified experimentally by Fromm.[2] The gain in performance is particularly noticeable when the IF is much greater than the pulse's spectral spread (see, for example, $\lambda=10$, Fig. 4). Too much noise is then passed by the IF along with the signal, and accordingly narrowing the video helps to eliminate some of the added noise products following rectification. The maximum effect occurs for videos slightly broader than the pulse spectrum. Of course, too pronounced a decrease in video width will eventually cut down the maximum signal level more rapidly than the noise, and the system will lose in effectiveness. Such behavior is observed in Figs. 2–4 as $\Omega \to 0$. On the other hand, for an IF narrower than the pulse (say, $\lambda=0.1$, Fig. 2) no improvement is observed: an infinitely wide video still gives the best results. This is because now the maximum signal strength is decreased more rapidly by the insertion of a video filter than is the r.m.s. noise level.

Figure 5(a) shows how the input power required for a given output ratio ($s_v/n_v=0$ db) varies with the pulse or IF band width, for a variety of video responses. The narrower the video the smaller the value of $\bar{\lambda}(\equiv \omega_g/\omega_b)$ at which the mean input power $\bar{P}$ is a minimum. The explanation lies in the fact that for a fixed IF characteristic the longer pulses have fewer significant components, among which correspondingly more of the pulse's energy is distributed; the narrow videos then remove relatively more noise than signal power in such circumstances, since a larger proportion of pulse energy lies in the low frequency harmonics passed by the video filter. The minimum power required for a given output (here $s_v/n_v=0$ db) occurs at match ($\lambda=1$), as predicted earlier. We note, moreover, that pulse shape is not a very critical factor, except when the video filters are very narrow compared to the pulse. Indeed, according to our "simple" criterion, performance at match (for a fixed PRF) is independent of pulse shape; (actually, there is a small modification because visibility of the pulse is somewhat affected by its widths).[3] Long pulses are no easier to detect than short ones in their respective matched conditions, since we observe merely the net increment in signal, which is quite independent of its duration. This is not true, of course, away from match, as Figs. 2 and 4 show: long pulses are then more easily detected than short ones.

Finally, we call attention to the phenomenon of *modulation suppression*, which is particularly noticeable when the signal is comparable to, and less than, the noise, and is characteristic of all nonlinear devices such as the second detector of our present receiver. For ex-

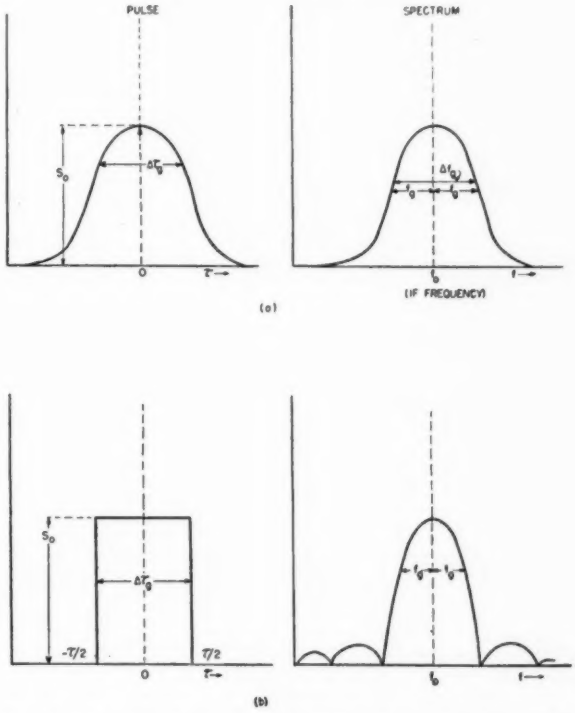


FIG. 1. The wave form and spectrum of (a), a Gaussian pulse, and (b), a rectangular pulse.

[4] This is easily deduced from Eq. (2.16) when $\lambda=1$; only the term $n=0$ in the expression for the noise Σ_n (Ω') (Eq. (2.8)) appears when the detector is a square-law device. We observe that

$$s_v(0)_G/n_v \doteq \sigma_0^2[(\Omega^4+2\Omega^2)/(\Omega^4+2\Omega^2+1)]^{\frac{1}{2}}, \quad (\lambda=1),$$

and clearly the coefficient of σ_0^2 (=a constant) is a maximum when $\Omega \to \infty$, i.e., $\omega_v \to \infty$. This result that the optimum video is one that is indefinitely wide was originally shown by Uhlenbeck (reference 3); see also footnote 19, reference 1, and the footnote following reference 3 above.

ample, when this second detector is the half-wave linear rectifier assumed here, an incoming signal that is weak relative to the noise appears quadratically[5] in the output, instead of linearly; suppression is also noted when the noise is weak, for then the carrier in its turn overwhelms the noise in nonlinear fashion until for sufficiently intense signals the noise becomes relatively ignorable and independent of carrier strength. The envelope, or modulation, is reproduced without distortion. Suppression is evident in Figs. 2–4, where for small signals the slope of the curves is 2, changing to a slope of unity as the input power is increased. Experiment[2] is in good agreement on this point.

II. ANALYSIS

The present theory follows directly from reference 1; we assume explicitly that all filter characteristics are Gaussian. Such responses, while not actually realizeable in practice, are analytically very convenient, and are nevertheless good approximations in most instances, as long as we are not too concerned with frequencies well away from resonance. Two different pulse shapes are considered here: (a) Gaussian and (b) rectangular (see Figs. 1a, 1b), the latter being closer to the particular waveform used in Fromm's work,[2] while the former is a more general shape; pulse shape itself, however, is not a very critical factor in most cases.

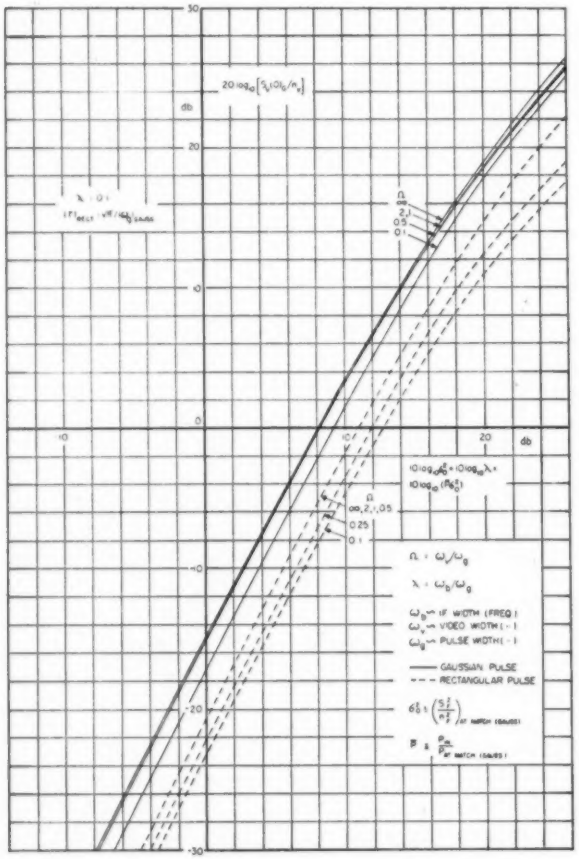


FIG. 2. Output video signal-to-noise ratios for unmatched pulse and IF ($\lambda<1$); same pulse power and peak amplitudes for Gaussian and rectangular pulses entering the IF.

Let

$$G_v(\omega)=|G_v(\omega)|e^{-i\omega t_0}, \quad \omega=2\pi f, \quad f>0 \qquad (2.1)$$

be the amplitude response of the inserted video filter; we assume here, and in all our various filters, a phase shift that is linear with frequency—i.e., no phase distortion. The mean-square output noise voltage after rectification is represented by

$$n_v^2=\int_0^\infty W_0(f)df=4\int_0^\infty df\int_0^\infty R_0(t)\cos\omega t dt. \qquad (2.2)$$

Here $W_0(f)$ and $R_0(t)$ are, respectively, the low frequency power spectrum and correlation function after detection, on the assumption of an infinite, uniform video response. If we take account of the finite characteristic of our inserted video filter, (2.2) is modified at once to give

$$n_v^2=\int_0^\infty W_0(f)|G_v(\omega)|^2df=\frac{2}{\pi}\int_0^\infty |G_v(\omega)|^2d\omega \times\int_0^\infty R_{0(n\times n),\,p=0}\cos\omega t dt. \qquad (2.3)$$

Since the pulsed signal is on only a small fraction of the repetition period, the signal×noise ($s\times n$) products generated in rectification are negligible, and the noise ×noise ($n\times n$) contributions alone are significant; furthermore, the ($n\times n$) noise is essentially unaffected by the presence of such a signal, so that we can safely replace $R_0(t)$ in (2.2) above by $R_0(t)_{(n\times n),\,p=0}$, which is the output correlation function for rectified noise without a signal ($p=0$). The $d-c$ term, corresponding to the constant part of $R_0(t)$, is excluded, since our video filters do not pass the steady component of the detected wave.

For the effect of the video filter on the signal we may write

$$s_v(t)_G=\frac{1}{2\pi}\int_{-\infty}^\infty e^{i\omega t}S_v(\omega)G_v(\omega)d\omega, \qquad (2.4)$$

where $S_v(\omega)$ is the Fourier transform of the unfiltered video output, given by $[s_v(t)_{p>0}-s_v(t)_{p=0}]$. For real signals $G_v(\omega)$ is required to be equal to $G_v(-\omega)^*$; $s_v(t)_{p>0}$ and $s_v(t)_{p=0}$ correspond respectively to the r.m.s. increments in video signal (infinite, uniform video filter) when there is a pulse and noise and when there is noise alone. The difference of these two quantities is consequently the mean net increment in the video level due

[5] In fact, it has been shown that all half-wave second detectors behave like a quadratic rectifier when the signal is weak compared to the noise. For details see D. Middleton, *Rectification of a Sinusoidally Modulated Carrier in Random Noise*, Proc. I.R.E. **36**, 1467 (1948), and Cruft Laboratory Technical Report No. 45, July 1, 1948.

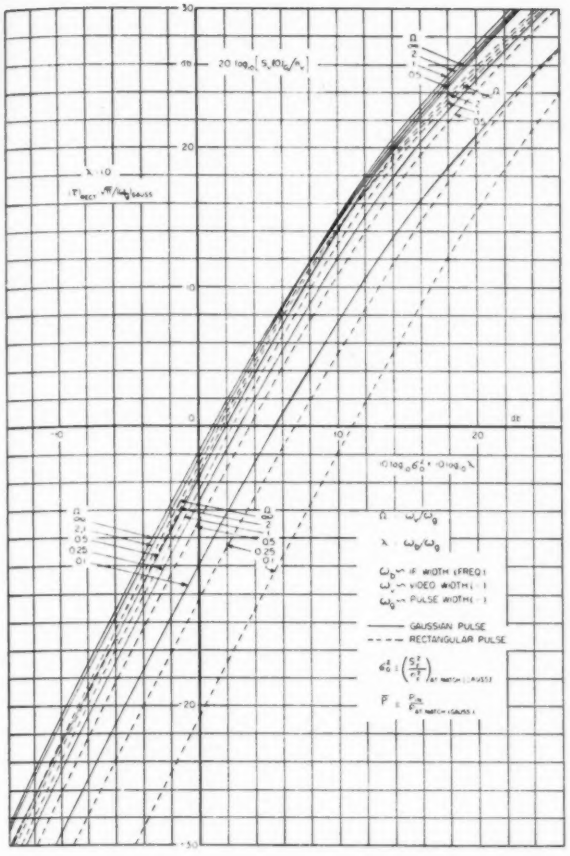


FIG. 3. Output video signal-to-noise ratios for matched pulse and IF; equal pulse power and peak amplitudes for Gaussian and rectangular pulses entering the IF.

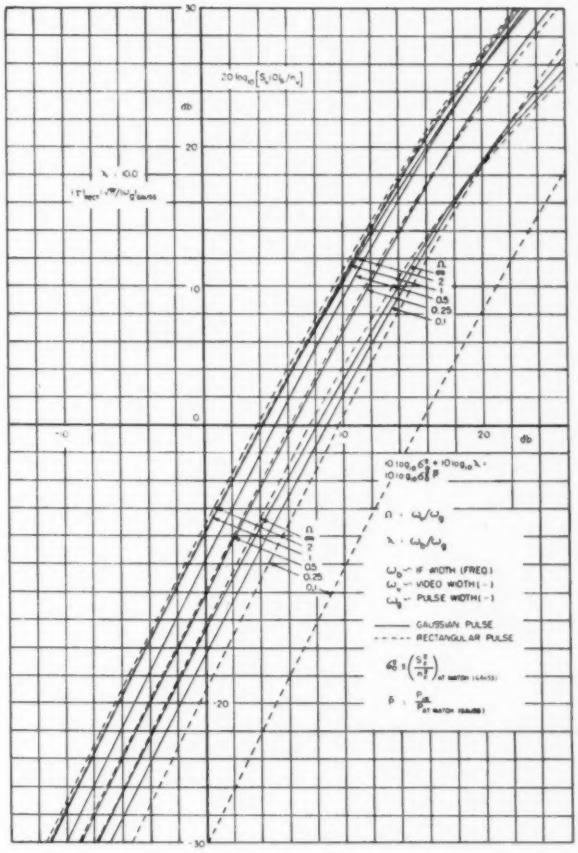


FIG. 4. Output video signal-to-noise ratios for unmatched pulse and IF ($\lambda>1$); same pulse power and peak amplitudes for Gaussian and rectangular pulses entering IF.

to the presence of a signal, corresponding to our definition in Section 1. (The notation here and in what follows is that of reference 1.) With the aid of (2.1) in (2.4) we obtain $s_v(t)_G$, observing that according to our "simple" criterion of visibility we want now the maximum value of $s_v(t)_G$; this corresponds to the maximum of the r.m.s. increment due to signal and noise over the surrounding noise background and is found by setting $t=t_0$, i.e., by selecting our origin on the time scale so that $t=t_0$ gives a maximum, a gating operation performed automatically by the eye. We have, finally,

$$s_v(0)_G=[s_v(t)_G]_{\max}=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega\,|G_v(\omega)|$$
$$\times\int_{-\infty}^{\infty}[s_v(t')_{p>0}-s_v(t')_{p=0}]e^{-i\omega t'}dt'. \quad (2.5)$$

Let us now calculate some specific examples. We choose $G_F(\omega)$ and $G_v(\omega)$ to be respectively the amplitude filter responses of the IF and video stages of our receiver and represent their moduli by

$$|G_F(\omega)|=e^{-\omega^2/2\omega_b^2}, \quad \omega_b=2\pi f_b,$$
$$|G_v(\omega)|=e^{-\omega^2/2\omega_v^2}, \quad \omega_v=2\pi f_v. \quad (2.6)$$

In both the above instances we have assumed a linear phase dependence with frequency. The various quantities f_b, f_v, etc., are described in Table I.

The low frequency correlation function for the noise following a half-wave linear detector is found to be [cf. Eq. (24), reference 1] exclusive of the steady component,

$$R_0(t)_{(n\times n),\,p=0}=\frac{\beta^2\psi}{8\pi}\sum_{n=0}^{\infty}\mu_n{}^2r_0(t)^{2n+2} \quad (2.7)$$

where

$$\mu_n=\frac{(2n)!}{2^{2n}n!(n+1)!}=\frac{(1/2)_n}{(2)_n},$$

$$r_0(t)=\exp(-\omega_b{}^2t^2/4); \quad \psi=n_F{}^2=\frac{W_0\omega_b}{2\pi^{\frac{1}{2}}}.$$

The quantities W_0 and ψ are the maximum spectral intensity of the noise entering the IF and the mean noise power output of the IF, respectively; $r_0(t)$ is therefore the (normalized) correlation function of the random noise leaving the IF stage; β is a constant (e.g., the dynamic transconductance) of the detector.

Substituting (2.1) and (2.7) into (2.3) and performing the integration yield the mean-square video noise

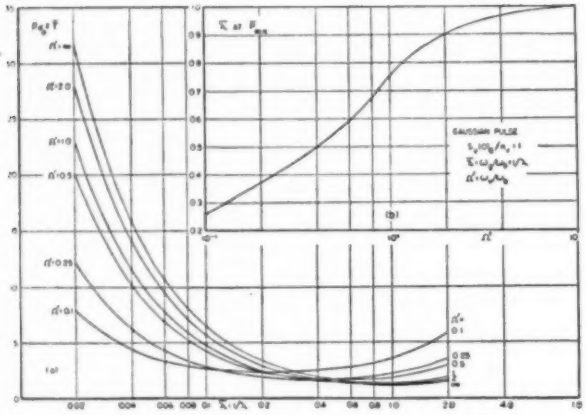


FIG. 5. (a) Average pulse power as a function of pulse length or IF width; (b) minimum average pulse power as a function of the video filter width.

voltage

$$n_v^2=\frac{\beta^2\psi}{8\pi}\sum_n(\Omega')$$

$$=\frac{\beta^2\psi}{8\pi}\sum_{n=0}^{\infty}\mu_n^2\frac{1}{(1+2(n+1)\omega_b^2/\omega_v^2)^{\frac{1}{2}}};$$

$$\Omega'\equiv\omega_v/\omega_b. \tag{2.8}$$

Observe that when the video is infinitely wide, i.e., $\omega_v\to\infty$, the series can be summed to give the familiar value $n_v^2=0.0435\beta^2\psi$, which is our earlier result, Eq. (25), reference 1. Note also that for vanishingly narrow videos ($\omega_v\to 0$) the output vanishes, as one would expect.

For the half-wave linear rectifier considered here, the increment in video signal output is[1,6]

$$s_v(t')_{p>0}-s_v(t')_{p=0}$$

$$=\frac{\beta\psi^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}[{}_1F_1(-\tfrac{1}{2};1;-s_F(t')^2/\psi)-1], \tag{2.9}$$

where ${}_1F_1$ is a confluent hypergeometric function and $s_F(t')$ is the r.m.s. envelope of the carrier after passage through the IF filter. In terms of the original r.m.s. modulation $s_I(t)$ entering the receiver we obtain for $s_F(t')$

$$s_F(t')=\frac{1}{2\pi}\int_{-\infty}^{\infty}G_F(\omega)e^{i\omega t'}d\omega\int_{-\infty}^{\infty}s_I(t)e^{-i\omega t}dt. \tag{2.10}$$

Before considering particular examples we remark that (2.9) is more readily handled if we approximate the hypergeometric function by $x/2-\gamma x^{\frac{1}{2}}$; ($\gamma=0.0763$, $0\le x\le 10$) in the manner of reference 1 (p. 962).

Case I. Gaussian Pulse

Here we choose

$$\sqrt{2}s_I(t)=S_0\exp(-\omega_g^2t^2/2), \tag{2.11}$$

where $S_0\equiv$peak value of the envelope and $\omega_g=2\pi f_g$. Then from (2.10) and (2.6) we obtain[7]

$$s_F(t')=s_F\exp(-t'^2b^2/4);\quad b\equiv\sqrt{2}\omega_b\omega_g/(\omega_b^2+\omega_g^2)^{\frac{1}{2}};$$
$$s_F=S_0b/2\omega_g. \tag{2.12}$$

Substituting our results (2.12) into (2.9), with the aid of the approximation for ${}_1F_1$, and applying this to (2.5) enables us to write finally for the maximum average signal

$$s_v(0)_G=\frac{\beta\psi^{\frac{1}{2}}p}{2(2\pi)^{\frac{1}{2}}(1+b^2/\omega_v^2)^{\frac{1}{2}}}$$
$$\times\{1-\gamma'p^{\frac{1}{2}}(1+b^2/\omega_v^2)^{\frac{1}{2}}/(\tfrac{2}{3}+b^2/\omega_v^2)^{\frac{1}{2}}\};$$
$$\gamma'=\gamma(8/3)^{\frac{1}{2}}=0.1246,\quad p\le 10. \tag{2.13}$$

The quantity p is defined as the mean-square maximum signal output of the IF divided by the mean-square noise amplitude, *viz.*: $p\equiv s_F(t)_{\max.}^2/n_F^2=s_F(0)^2/n_F^2=s_F^2/\psi$. Again we may check (2.13) by noting that as the video is made very wide our signal $s_v(0)_G$ approaches the earlier form, Eq. (25) of reference 1. The desired signal-to-noise ratio $s_v(0)_G/n_v$ follows at once from (2.13) and (2.8). We see from (2.13) that for weak signals the output signal power is proportional to the *square* of the input carrier-to-noise ratio p, an example of modulation suppression discussed in the preceding section. On the other hand, strong signals suppress the noise, so that when $p^2\gg 1$ we obtain from the asymptotic development of ${}_1F_1$ (see Eq. (A3.3) of reference 6) the average video increment

$$s_v(0)_G\backsimeq\frac{\sqrt{2}}{\pi}\beta(\psi p)^{\frac{1}{2}}(1+b^2/2\omega_v^2)^{-\frac{1}{2}},\quad p>20, \tag{2.14}$$

which is now directly proportional to $(p)^{\frac{1}{2}}$. The two expressions (2.13) and (2.14) for the video signal enable us to cover the entire range of values of $s_v(0)_G/n_v$ *vs.* p with a small region of interpolation for $10<p<20$.

At this point it is convenient to introduce a number of parameters

$$\lambda=\omega_b/\omega_g;\quad \Omega=\omega_v/\omega_g;\quad \Omega'=\omega_v/\omega_b=\Omega/\lambda=\Omega\bar{\lambda}. \tag{2.15}$$

The signal-to-noise ratio becomes finally [see (2.8)]

$$\left[\frac{s_v(0)_G}{n_v}\right]_{\text{Gauss}}=\frac{p(1-0.1246(p)^{\frac{1}{2}}[1+2\lambda^2/\Omega^2(1+\lambda^2)]^{\frac{1}{2}}/[\frac{2}{3}+2\lambda^2/\Omega^2(1+\lambda^2)]^{\frac{1}{2}})}{[1+2\lambda^2/\Omega^2(1+\lambda^2)]^{\frac{1}{2}}\{\sum_n(\Omega')\}^{\frac{1}{2}}},\quad p\le 10, \tag{2.16}$$

[6] D. Middleton, "Some general results in the theory of noise through nonlinear devices," Quart. App. Math. **5**, 445 (1948), Section 4 and Appendix A3.

[7] There is a misprint in the expressions for s_F in Eq. (61b) of reference 1; $\sqrt{2}$ should be replaced by 2.

TABLE I.

Pulse	Amplitude of wave form	Power spectrum	Time width* $(\Delta t)_g$	Spectral width* $(\Delta f)_g$	s_F^2/n_F^2	P_{in}
1. Gauss	$S_0\exp(-\omega_g^2t^2/2)$	$(S_0^2\pi/2\omega_g)\exp(-\omega^2/\omega_g^2)$	$\frac{(\log_e 2)^{\frac{1}{2}}}{\pi f_g}=0.265/f_g$	$2(\log_e 2)^{\frac{1}{2}}f_g=1.666\cdot f_g$	$\frac{(\pi)^{\frac{1}{2}}S_0^2\omega_b}{W_0(\omega_g^2+\omega_b^2)}$	$\frac{S_0^2(\pi)^{\frac{1}{2}}}{2\omega_g T}$
2. Rectangular	S_0, $-\tau/2<t<\tau/2$; 0, $\lvert t\rvert>\lvert\tau/2\rvert$	$4S_0^2\frac{\sin^2(\omega\tau/2)}{\omega^2}$	τ	$2.782/\pi\tau=0.866/\tau$	$\frac{S_0^2\tau}{2T}$	$(\pi)^{\frac{1}{2}}S_0^2\frac{\Theta^2\left(\frac{\tau\omega_b}{2\sqrt{2}}\right)}{W_0\omega_b}$
Filter						
3. IF (centered about IF freq. f_0)		$\exp(-\omega^2/\omega_b^2)$		$(\Delta f)_F=2(\log_e 2)^{\frac{1}{2}}f_b$		
4. Video		$\exp(-\omega^2/\omega_v^2)$		$(\Delta f)_v=(\log_e 2)^{\frac{1}{2}}f_v$		

* The widths of the filters and pulses are measured between half-power (or intensity) points—i.e., points at which the intensity is one-half its maximum value.

Ω = [video filter width (freq.)]/[½ pulse width (freq.)]; $\Omega'=\Omega/\lambda=\Omega\bar{\lambda}$.
λ = [IF filter width (freq.)]/[pulse width (freq.)] $=\bar{\lambda}^{-1}$; T = PRF; $\tau\ll T$.

and for strong signals

$$\left[\frac{s_v(0)_G}{n_v}\right]_{\text{Gauss}}\backsimeq 4(p/\pi)^{\frac{1}{2}}\left\{\left[1+\frac{\lambda^2}{\Omega^2(1+\lambda^2)}\right]\left[\sum_n(\Omega')\right]\right\}^{-\frac{1}{2}},\qquad p>20. \quad (2.17)$$

Figure 3 illustrates $s_v(0)_G/n_v$ when $s_F/n_F=(p)^{\frac{1}{2}}$ is chosen as independent variable in the matched condition $\lambda=1(\omega_b=\omega_g)$. The effect of the different videos is observed through the quantity Ω. When the pulse and IF filter are not matched ($\lambda\neq 1$) we may represent the ratio p by

$$p=[2\lambda^2/(1+\lambda^2)]\sigma_0^2, \quad (2.18)$$

where σ_0^2 is taken to be $(s_F/n_F)^2$ at match. In terms of the input power to the IF, which is simply the (energy/pulse)×(number of pulses/sec.), we may also write [cf., Eq. (68), reference 1),

$$p=[2\lambda/(1+\lambda^2)]\sigma_0^2\bar{P}\bar{T}, \quad (2.19)$$

in which $\bar{P}=P_{in}/P_0$ and $\bar{T}=T_{in}/T_0$, the subscript (0) referring to the input power to the IF (P_0) and pulse repetition period (T_0) at match, while P_{in} and T_{in} are these quantities for arbitrary conditions of operation, not necessarily at match. Figures 2 and 4 show $s_v(0)/n_v$ at mismatch, and here we have chosen as our independent variable $\sigma_0^2\lambda$, or $\sigma_0^2\bar{P}\bar{T}$, in order to take into account the added power input necessary to achieve a given output when away from match.

We mention briefly two additional limiting cases: for a fixed pulse width and very wide IF ($\omega_b\to\infty$, $\lambda\to\infty$; $\Omega'\to 0$, $\Omega=\Omega$) s_v/n_v vanishes as λ^{-1}, since $\sigma_0^2\sim\omega_b^{-1}$, whereas for a fixed IF and a very narrow pulse (in time) ($\omega_g\to\infty$, $\lambda\to 0$, $\Omega'=\Omega'$; $\Omega\to 0$) we find at once that $s_v/n_v\sim\lambda^2$. Thus in both instances the output vanishes; in the former because of the added noise, in the latter because of vanishing signal strength. Similar remarks apply for the rectangular pulse treated below.

Case II. Rectangular Pulse

Here the shape of the signal wave entering the IF is

$$\sqrt{2}s_I(t)=S_0,\quad -\tau/2<t<\tau/2;\quad =0,\quad \lvert t\rvert>\lvert\tau/2\rvert. \quad (2.20)$$

S_0 is again the peak amplitude, and τ measures the duration of the pulse between half-power points. The various IF and video filters are Gaussian as before (cf. Eq. (2.6)]. The envelope of the pulse leaving the IF follows immediately from (2.10) and the result

$$\int_0^\infty\frac{\sin ax}{x}\exp(-bx^2/2)dx=\pi\phi^{(-1)}(a/b^{\frac{1}{2}})=\frac{\pi}{2}\Theta[a/(2b^{\frac{1}{2}})], \quad (2.21a)$$

where

$$\Theta(x)=\frac{2}{\pi^{\frac{1}{2}}}\int_0^x e^{-t^2}dt \quad (2.21b)$$

is the well-known error function. We have finally

$$s_F(t')=s_F\{\Theta([t'+\tau/2]\omega_b/\sqrt{2})-\Theta([t'-\tau/2]\omega_b/\sqrt{2})\}/2\Theta(\tau\omega_b/2\sqrt{2}), \quad (2.22)$$

in which

$$s_F=S_0\Theta\left(\frac{\tau\omega_b}{2\sqrt{2}}\right)\Big/\sqrt{2}. \quad (2.22a)$$

In terms of our definition for p we see that

$$p_{\text{rect.}}=(\pi)^{\frac{1}{2}}S_0^2\Theta(\tau\omega_b/2\sqrt{2})^2/W_0\omega_b=2\Theta\left(\frac{\tau\omega_b}{2\sqrt{2}}\right)^2\sigma_0^2. \quad (2.22b)$$

Compare this with Eq. (2.18) which gives p_{Gauss}; if we make such a comparison on the basis of equal energy/pulse into the IF, i.e.,

$$E_{\text{Gauss}}=E_{\text{rect.}}=\int_{-\infty}^{\infty}s_I(t)^2dt,$$

the relation between τ and ω_g is found from (2.11) and (2.20) to be

$$\tau=(\pi)^{\frac{1}{2}}/\omega_g. \tag{2.23}$$

Then we have

$$p_{\text{rect.}}=\left(\frac{1+\lambda^2}{\lambda^2}\right)\Theta\left(\frac{\pi^{\frac{1}{2}}\lambda}{2\sqrt{2}}\right)^2 p_{\text{Gauss}}, \quad (\tau=\pi^{\frac{1}{2}}/\omega_g). \tag{2.24}$$

Thus we are able to discuss the case of rectangular pulses in terms of the parameters (2.15) of the wholly Gaussian system.

With (2.22) for $s_F(t')$ we now use our approximation for ${}_1F_1$ but omit the term containing $s_F(t')^3$, as it leads to analytical difficulties out of proportion to the results achieved; our final expression is not seriously affected, since with the aid of the asymptotic development for ${}_1F_1$ we can interpolate with little error in the region not covered by the small-signal result involving $s_F(t')^2$ alone ($p<1$) and so obtain s_v/n_v as a function of $p^{\frac{1}{2}}$ (or $\sigma_0(\lambda)^{\frac{1}{2}}$) over the entire range of these variables; the parallel behavior with the Gaussian pulse aids the interpolation.

The increment in video signal (2.9) is accordingly[8]

$$s_v(0)_G\doteq p\frac{\beta\psi^{\frac{1}{2}}\omega_v}{16\pi}\int_{-\infty}^{\infty}dt'e^{-\omega_v^2t'^2/2}\{\Theta[(t'+\tau/2)\omega_b/\sqrt{2}]$$

$$-\Theta[(t'-\tau/2)\omega_b/\sqrt{2}]\}^2/\Theta^2\left(\frac{\tau\omega_b}{2\sqrt{2}}\right), \quad p<\tfrac{1}{2}. \tag{2.25}$$

The integrand in (2.25) may be evaluated in straightforward manner if we use a contour integral representation of the error function and the related $\phi^{(j)}(x)$, which are specifically

$$\phi^{(j)}(x)=\frac{d^{(j)}}{dx^{(j)}}\frac{e^{-x^2/2}}{(2\pi)^{\frac{1}{2}}}, \quad j=0, 1, \cdots,$$

and are related to the Hermitian polynomials $H_j(x)$ by $(-1)^jH_j(x)e^{-x^2/2}/(2\pi)^{\frac{1}{2}}=\phi^{(j)}(x)$; the $\phi^{(j)}$ are tabulated up to $j=6$ in T. C. Fry's *Probability and Its Engineering Uses* (D. Van Nostrand Company, Inc., New York, 1928), p. 456.

We obtain finally for the video increment[9]

$$s_v(0)_G\doteq\frac{p\beta\psi^{\frac{1}{2}}}{2(2\pi)^{\frac{1}{2}}}\sum_{n=0}^{\infty}\frac{4}{(2n)!(\Omega'^2+1)^{2n}}$$

$$\times\phi^{(2n-1)}\left(\frac{\tau\omega_v}{2(\Omega'^2+1)^{\frac{1}{2}}}\right)^2\Big/\Theta\left(\frac{\tau\omega_b}{2\sqrt{2}}\right)^2, \quad p\leq\tfrac{1}{2}. \tag{2.26}$$

Compare this with the leading term of (2.13) for the Gaussian pulse. Notice that for an infinitely wide video response ($\omega_v\to\infty$), $s_v(0)_G$ reduces to $s_F^2\beta/2(2\pi\psi)^{\frac{1}{2}}$, as we would expect. Pulse shape is no longer significant relative to the video; all harmonics of the envelope are passed when $\omega_v\to\infty$ and only the energy/pulse is important.

The output signal-to-noise ratio for the rectangular pulse is obtained from (2.8) and (2.26) and is

$$\left[\frac{s_v(0)_G}{n_v}\right]_{\text{rect.}}\doteq p\left\{\sum_{n=0}^{\infty}4\phi^{(2n-1)}\left(\frac{\tau\omega_v}{2(\Omega'^2+1)^{\frac{1}{2}}}\right)^2\right.$$

$$\left.\div\Theta^2\left(\frac{\tau\omega_b}{2\sqrt{2}}\right)(2n)!(\Omega'^2+1)^{2n}\right\}\Big/(\textstyle\sum_n(\Omega'))^{\frac{1}{2}},$$

$$p<\tfrac{1}{2}. \tag{2.27}$$

The noise term is unchanged, since only ($n\times n$) products ($p=0$) enter effectively, and the amount of noise power available in the low frequency spectra region after our linear rectifier then depends on IF filter shape alone, cf. (2.7) and (2.8). Again with the help of the contour integral representation of the error function and $\phi^{(j)}(x)$, we find that for large carrier strengths the video signal is

$$[s_v(0)_G]_{\text{rect.}}\backsimeq\frac{\sqrt{2}\beta s_F}{\pi}\Theta\left(\frac{\tau\omega_v}{2(2\Omega'^2+2)^{\frac{1}{2}}}\right)\Big/\Theta\left(\frac{\tau\omega_b}{2\sqrt{2}}\right);$$

$$s_F^2/\psi>20, \tag{2.28}$$

analogous to (2.14). The signal-to-noise ratio is therefore

$$[s_v(0)_G/n_v]_{\text{rect.}}$$

$$\backsimeq 4(p/\pi)^{\frac{1}{2}}\Theta\left(\frac{\tau\omega_v}{2(2\Omega'^2+2)^{\frac{1}{2}}}\right)\Big/\Theta\left(\frac{\tau\omega_b}{2\sqrt{2}}\right)(\textstyle\sum_n(\Omega'))^{\frac{1}{2}},$$

$$p>20, \tag{2.29}$$

a result to be contrasted with (2.17). Figures 2–4 also give (2.27) and (2.29).

ACKNOWLEDGMENT

The author wishes to express his appreciation to Miss Marilyn Lang and Mrs. Roger Stokey, who performed most of the computations in this paper.

[8] p is written p_{Gauss} or $p_{\text{rect.}}$ only when it s not clear from the context which is meant.

[9] The details of the evaluation are given in Technical Report No. 57, Cruft Laboratory, Harvard University (August 10, 1948).

Negative Wire Corona at High Temperature and Pressure

L. R. Koller and H. A. Fremont
Research Laboratory, General Electric Company, Schenectady, New York
(Received December 27, 1949)

The corona characteristics of air and methyl chloride were studied in a modified type of Cottrell precipitator at temperatures from room temperature to 500°C and pressures from 1 to 5 atmospheres. Within this range the corona current is a power function of voltage with gas density as a parameter. This power is independent of density. Values are given for air and methyl chloride.

INTRODUCTION

In the course of the development of a high temperature, high pressure electrostatic precipitator, it was necessary to determine the corona characteristics of air and methyl chloride. These studies were made in a modified type of Cottrell precipitator at temperatures from room temperature to 500°C and pressures from 1 to 5 atmospheres. The wire sizes were varied from 0.005″ to 0.02″ in diameter. Negative wire corona only was investigated.

APPARATUS

A drawing of the precipitator used is shown in Fig. 1. It was originally constructed for the removal of finely divided powder from a process gas stream. It consisted essentially of a wire mounted along the axis of a 2″ diameter tube. The latter was of seamless steel 36″ long and 2″ in diameter. Great care was used in the construction of the unit to remove sharp edges and roughened surfaces.

A drawn tungsten wire was used for the central electrode. It was not given any cleaning or polishing treatment. No significant pitting or corrosion of the wire was observed during the tests.

The wire was mounted in a bayonet-like fixture at the bottom and spring loaded by means of a tungsten spring at the top. The ends were fastened by ball and socket joints formed by fusing the ends of the wire to small balls about 50 mils in diameter. This eliminated any sharp projecting ends of wire. The purpose of the bayonet fixture was to leave the bottom part of the equipment free for both visual observation of the wire during operation, and also to simplify removal of any powder precipitated when the unit was used as a precipitator.

Zirconium silicate insulators similar to submarine insulators were supplied by the Locke Insulator Company. The leakage over these insulators was negligible in the range of currents investigated.

The unit was wrapped with nichrome wire and insulated with pipe lagging so that it could be heated electrically. Temperatures were determined by three chromel alumel thermocouples peened into the metal and equally spaced along the 2″ pipe.

The electrical circuit is shown in Fig. 2. It consisted of a voltage doubler capable of supplying 25×10^{-3} amps at 40 kv. The positive terminal of the power supply was connected to the precipitator shell which was grounded, and the negative side was connected to the wire. The voltage was controlled by means of a variac in the primary of the high voltage transformer. Voltage (average) was measured on a voltmeter consisting of a microammeter with a 30 megohm series resistance. In order to limit the current in case of arc over an 8013A kenetron, with its emission limited by operating the filament at reduced temperature was included in the high potential side of the line.

The precipitator was assembled with piping containing a flow indicator, pressure gauge and throttle valve. With this arrangement it was possible to introduce various gases at pressures up to 5 atmospheres and make measurements with a small steady flow of gas through the tube. The gas flow was introduced at the upper end of the precipitator as shown in Fig. 1.

PROCEDURE AND PRELIMINARY TESTS

Preliminary tests were made in air at 1 atmosphere. For these tests the end of the unit was sealed with a

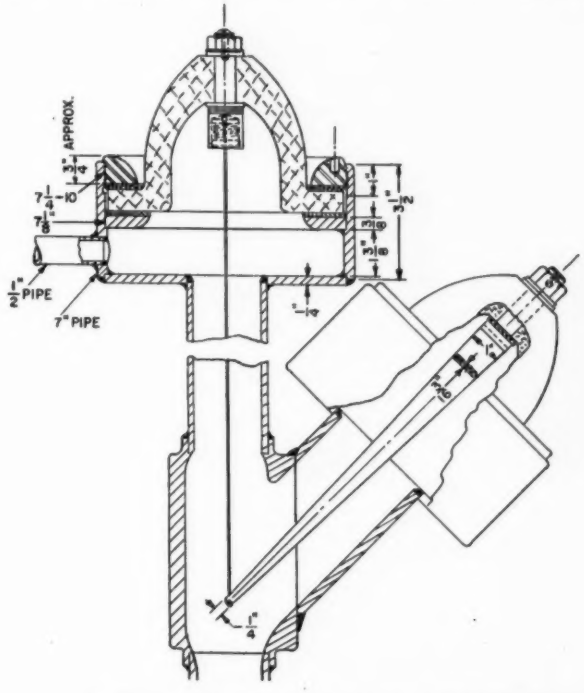


FIG. 1. Design of electrostatic precipitator.

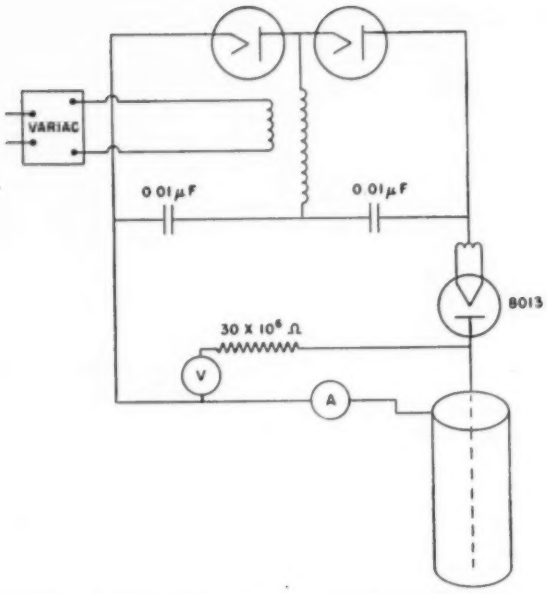


FIG. 2. Electrical circuit for corona measurements.

sheet of $\frac{1}{8}$" transparent cellulose acetate. This made it possible to observe the corona visually by looking in through the end of the tube along the wire.

Voltage was applied to the precipitator and gradually increased while the wire was observed through the open end of the tube. Corona was first noticed around 6000 volts and increased in intensity with increasing voltage. It appeared to be quite uniform along the wire, and there was no tendency to spark or arc over at the ends of the wire or over the insulators. As the voltage was increased still further, the wire was seen to vibrate. At 30 kv, corresponding to a current of about 10 ma spark-over generally took place near the middle of the wire. This may have been due to the close approach to the wall as a result of the vibration.

RESULTS

Typical volt ampere characteristics at 1 atmosphere and room temperature are shown in Fig. 3 for air, for a mixture of air and approximately 19 percent of CCl_4 by volume and an atmosphere of pure CH_3Cl. These curves show a shift toward higher voltage in going from air to CCl_4 to CH_3Cl as well as a decrease in slope. They illustrate the well-known effect of halogens in suppressing corona.

An extensive series of measurements was made first in air and later in CH_3Cl at pressures from 1 to 5 atmosphere and temperatures from room temperature to 300°C. The temperatures chosen were room temperature, 100°C, 200°C and 300°C. At each of these temperatures the volt ampere characteristics were measured at 1, 2, 3 and 4 atmospheres pressure with a small flow of gas through the tube. Figures 4 and 5 show the original observations. The curves at any one temperature show a shift towards higher voltages and a decrease in slope as the pressure is increased.

It was found that over the range of corona currents from 0.1×10^{-3} to 1×10^{-3} amp. the volt ampere characteristics gave straight lines when plotted on log-log paper. It seemed reasonable to assume that temperature and pressure had no effect on the discharge other than their effect on the density of the gas, that is the number of molecules between the electrodes. Thus the characteristics of the discharge should be uniquely determined by the mass of the gas between the electrodes.[1] Accordingly, at any given current the voltage should be a function of the gas density rather than of the temperature and pressure. To test this, values of voltage at a given current were plotted as a function of the density of the gas, taking the density at room temperature and 1 atmosphere as unity. The family of curves for air and methyl chloride are shown in Figs. 6 and 7, respectively. These figures show that at any given current the values of voltage plotted as a function of density fall on a smooth curve within the limits of the experimental error regardless of the temperature and pressure at which the measurements were made. The agreement is good except for the measurements at room temperature which depart considerably from the curve at the lower currents.

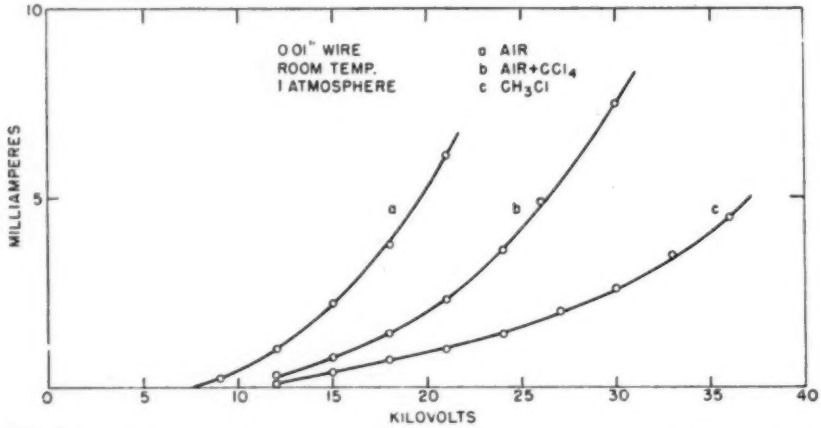


FIG. 3. Corona characteristics in various gases.

[1] Pauthenier and Rousse [Comptes Rendus **225**, 1293 (1947)], show that the critical gradient for the threshold of the corona effect is a function of the specific mass of air.

All of these measurements are in a range of voltages well above the corona starting voltage. The theoretical starting voltage calculated from the equation

$$v = 33.9 g r_i \log(r_a/r_i)[1-(0.295/g r_i)],^2$$

where v = corona starting voltage in kv, r_a = radius of outer cylinder in cm, r_i = radius of wire in cm, g = air density taking air at stated conditions = 1 is also shown in Fig. 7.

From the curves of Figs. 6 and 7 it is possible to read off the voltages as a function of current with gas density as a parameter. When plotted on log-log paper

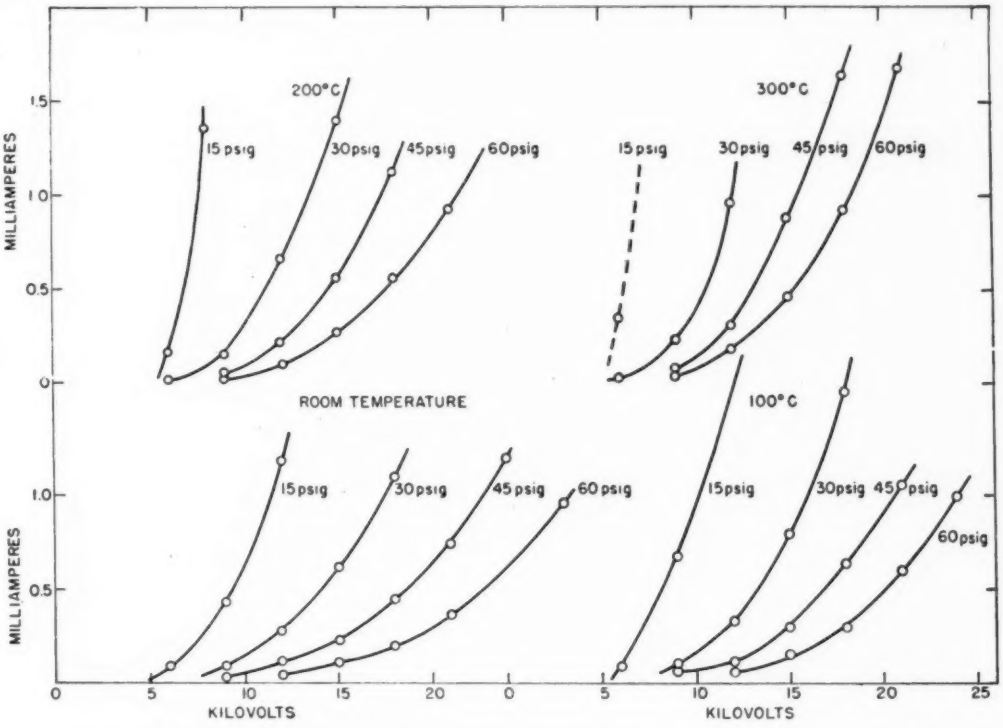


FIG. 4. Corona characteristics of 0.01" wire in air.

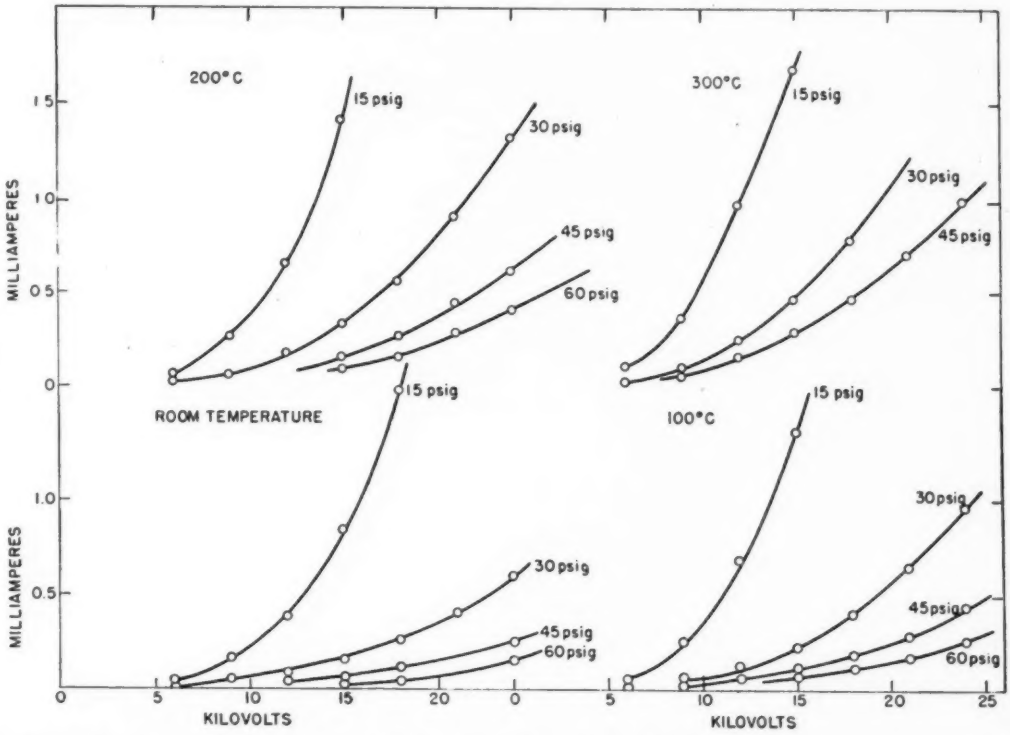


FIG. 5. Corona characteristics of 0.01" wire in methyl chloride.

[2] A. von Engel and M. Steenbeck, *Elektrische Gasentladungen* (J. Springer, Berlin, 1934), Vol. 2, p. 212.

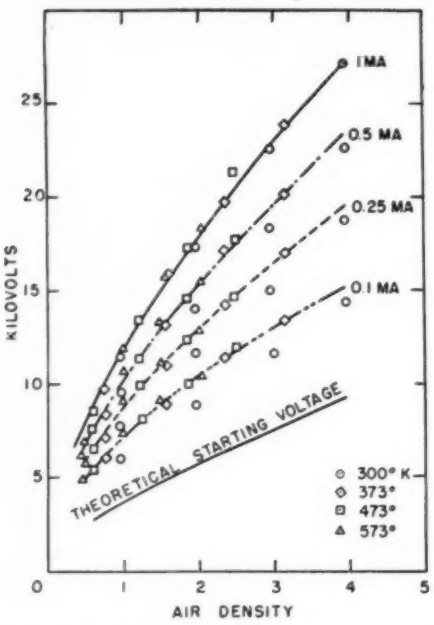


FIG. 6. Corona characteristics as a function of air density for 0.01″ wire.

TABLE I. Values of log A and K for air and methyl chloride.

d	Air	CH_3Cl	Air	CH_3Cl
	log A	log A	K	K
1	4.5	3.5	4.2	2.8
2	5.2	4.1	4.2	2.8
3	5.7	4.4	4.2	2.8
4	6.05	4.7	4.2	2.8

these volt ampere characteristics are straight lines, all having the same slope. They may be represented by an equation

$$\log i = \log A + K \log V$$

where i = current in milliamperes, V = voltage in kilovolts, A = a constant, K = a constant. They show that the current increases with the same power of the voltage at all values of density. This power is 4.2 for air and 2.8 for methyl chloride. The constants for different values of gas density are shown in Table I.

A series of measurements was carried out in air at various pressures at room temperature to determine the effect of wire size on corona characteristics. Measurements were made using tungsten wires 0.005″ and 0.02″ in diameter respectively to compare with measurements previously made with 0.01″ wire. These measurements showed that for currents of about 0.1×10^{-3} amps the effect of wire diameter on corona voltage was negligible. At 1×10^{-3} amp. the voltage increased about 3 kilovolts as the wire diameter was increased from 0.005″ to 0.02″. The gain in mechanical strength obtained by using a large diameter wire probably more than compensates for the small increase in voltage.

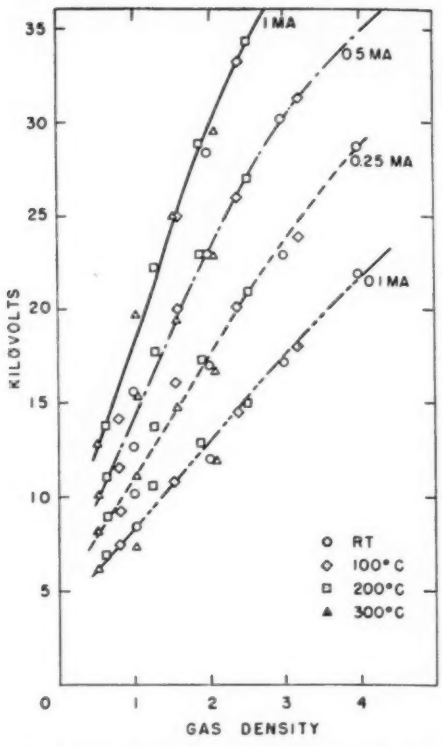


FIG. 7. Corona characteristics as a function of methyl chloride density for 0.01″ wire.

SUMMARY

The volt ampere characteristics of a corona discharge in a Cottrell precipitator are uniquely determined by the mass of gas between the electrodes and are independent of the temperature and pressure of the gas.

The corona current is a power function of the voltage. The power is independent of the gas density. Values of the power are given for air and methyl chloride.

Variations in corona characteristics with wire sizes are small at voltages above the corona starting voltage.

The Radiation from a Transverse Rectangular Slot in a Circular Cylinder*

SAMUEL SILVER AND WILLIAM K. SAUNDERS
Division of Electrical Engineering and the Antenna Laboratory, University of California, Berkeley, California
(Received January 27, 1950)

General results obtained previously for slots of arbitrary shape are applied to the case of a transverse rectangular slot in a circular cylinder. It is shown that the principal transverse plane pattern of such a slot in which the excitation has only a circumferential tangential electric field component is the same as the pattern generated by an infinite axial slot with the same circumferential excitation. Theoretical and experimental curves are given for the narrow-width half-wave-length slot.

I. INTRODUCTION

IN a previous paper[1] we developed general expressions for the external field produced by a slot of arbitrary shape in the wall of a circular cylindrical wave guide, on the assumption that the tangential electric field in the slot is a prescribed function. The present paper deals with the application of these results to the particularly simple configuration of a transverse rectangular slot: a slot that, on development of the cylinder into a plane, becomes a rectangle with boundaries respectively transverse and parallel to the axis of the cylinder. This configuration has been investigated by others by methods that differ from ours and are of less general applicability: C. H. Papas[2] has obtained expressions for the field for the case when the tangential electric field in the slot has only an axial component. G. Sinclair[3] has treated the case of a narrow rectangular slot, oriented arbitrarily on the surface of the cylinder, in which the tangential electric field is transverse to the slot axis.

The results given in this paper are restricted to the far-zone field of the slot; for this region the integrations can be performed and calculations can be made with relative ease. The computations have been made for the especially important case of a narrow slot having an axial extent of a half-wave-length. Experimental patterns have also been obtained. The data that are given here are for two values of cylinder radius that are likely to be used in microwave antenna designs.

II. THE FIELD DISTRIBUTION IN THE SLOT

The slot configuration and the various coordinates to be used are shown in Fig. 1. The slot is bounded axially by the planes $z=\pm l/2$ and in azimuth by the planes $\phi=\pm\phi_0$; or, in the notation used in I, we have for the boundaries of the slot

$$z_1=-l/2;\quad z_2=+l/2;\quad \phi_1(z)=-\phi_0;\quad \phi_2(z)=+\phi_0.$$

In formulating possible tangential electric field distributions in the slot we are guided by the consideration that at the boundaries of the slot the component of the distribution that is tangential to the boundary must be zero. In general, the excitation of the slot may be conceived as a superposition of many modes of field distribution each of which satisfies the fundamental requirement. This is analogous to the general excitation of a thin wire antenna which may be synthesized by superposition of characteristic sinusoidal distributions that satisfy the requirement that the current be zero at the ends.

The geometry of the configuration suggests that in each of the fundamental modes of excitation the field components are separable functions of ϕ and z; thus,

$$E_\phi(a,\phi,z)=f_1(\phi,z)=F_1(\phi)G_1(z), \tag{1a}$$

$$E_z(a,\phi,z)=f_2(\phi,z)=F_2(\phi)G_2(z). \tag{1b}$$

With the boundary condition at the edges of the slot in mind we are led to postulate that

$$G_1(z)=\begin{cases}\sin(p\pi z/l) & p, \text{ even integer},\\ \cos(p\pi z/l) & p, \text{ odd integer};\end{cases} \tag{2a}$$

$$F_2(\phi)=\begin{cases}\sin(p\pi\phi/2\phi_0) & p, \text{ even},\\ \cos(p\pi\phi/2\phi_0) & p, \text{ odd}.\end{cases} \tag{2b}$$

The functions $F_1(\phi)$ and $G_2(z)$ are not so clearly indicated. Certain forms are suggested for them by the point of view that the slot is a very shallow section of wave guide, in consequence of which the field configuration is like that over the cross section of a wave guide. On this basis we should have

$$F_1(\phi)=\begin{cases}\cos(q\pi\phi/2\phi_0) & q, \text{ even},\\ \sin(q\pi\phi/2\phi_0) & q, \text{ odd};\end{cases} \tag{3a}$$

$$G_2(z)=\begin{cases}\cos(q\pi z/l) & q, \text{ even},\\ \sin(q\pi z/l) & q, \text{ odd}.\end{cases} \tag{3b}$$

The limitations of the wave guide picture must be realized. It provides no basis for relating E_ϕ^{pq} to E_z^{pq} in the slot in the manner that they are related in the corresponding rectangular wave guide. Furthermore, there is some question as to whether the field components are bounded; infinities may exist at points on the slot

* Presented at the Navy Electronics Laboratory Antenna Conference, San Diego, April 1949 and at the West Coast Convention of the I.R.E. in September 1949.

[1] S. Silver and W. K. Saunders, J. App. Phys. **21**, 153 (1950); this paper will be designated hereafter as I.

[2] C. H. Papas, Cruft Laboratory Technical Report No. 61, Oct. 1948 (ONR Contract N5-ori-76); J. Math. and Phys. **28**, No. 4 January 1950.

[3] G. Sinclair, Ohio State University Antenna Laboratory Project Report 301-17, Sept. 1, 1949 (A.M.C. Contract W33-038 ac 16520 (17380)).

TABLE I. Values for which $H_0^{(2)\prime}(ka)/H_n^{(2)\prime}(ka) \leq 0.0001$ and $\sin(n\phi_0)/(n\phi_0) \geq 0.9$.

ka	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
n	5	6	8	9	9	10	11	12	13	13
ϕ_0 (rad.)	0.158	0.132	0.098	0.088	0.088	0.079	0.072	0.066	0.061	0.061

boundaries at which the field component is normal to the edge.[4]

III. THE FAR-ZONE FIELD

The general expressions for the far-zone field are given in the set of Eqs. (23) in I. In the present case, for a single mode of E_ϕ and E_z in the slot, the electric field components are

$$E_\theta = -\frac{1}{2\pi^2}\frac{e^{-jkR}}{R}\sum_{n=-\infty}^{\infty}\left\{\frac{(j^{n+1}e^{-jn\phi})}{\sin\theta H_n^{(2)}(ka\sin\theta)} \times \int_{-l/2}^{l/2} G_2(\xi)e^{jk\xi\cos\theta}d\xi \int_{-\phi_0}^{\phi_0} F_2(\beta)e^{jn\beta}d\beta\right\}, \quad (4a)$$

$$E_\phi = \frac{1}{2\pi^2}\frac{e^{-jkR}}{R}\sum_{n=-\infty}^{\infty}\left[\frac{j^n e^{-jn\phi}}{H_n^{(2)\prime}(ka\sin\theta)} \times \left\{\int_{-l/2}^{l/2} G_1(\xi)e^{jk\xi\cos\theta}d\xi \int_{-\phi_0}^{\phi_0} F_1(\beta)e^{jn\beta}d\beta + \frac{n\cot\theta}{ka\sin\theta}\int_{-l/2}^{l/2} G_2(\xi)e^{jk\xi\cos\theta}d\xi \times \int_{-\phi_0}^{\phi_0} F_2(\beta)e^{jn\beta}d\beta\right\}\right]. \quad (4b)$$

It will be observed that there is a striking difference between the fields produced by the respective modes of excitation in which there is no E_z-component in the slot and in which there is no E_ϕ-component in the slot. In the former case, when $F_2=G_2=0$, the radiation field is linearly polarized having only an E_ϕ-component. In the latter case, where $F_1=G_1=0$, the radiation field has both an E_ϕ- and E_z-component and, except in certain special planes, is elliptically polarized in general. These results can be correlated with the fact that the slot can be set into equivalence with a magnetic current sheet,[5] the direction of the current associated with a given tangential electric field component being transverse to that component.

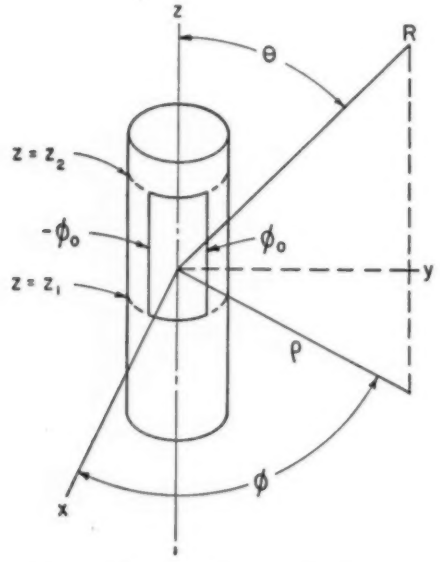


FIG. 1. The transverse rectangular slot.

There are further points of interest to be noted for the case where the tangential electric field in the slot has only an E_ϕ-component. The far-zone field is given by Eq. (4b) with F_2 and G_2 set equal to zero; on introduction of self-compensating factors of $\sin\theta$ in the appropriate places, we have

$$E_\phi = \frac{1}{2\pi^2}\frac{e^{-jkR}}{R}\int_{-l/2}^{l/2} G_1(\xi)\sin\theta e^{jk\xi\cos\theta}d\xi \times \sum_{n=-\infty}^{\infty}\frac{(j^n e^{-jn\phi})}{\sin\theta H_n^{(2)\prime}(ka\sin\theta)}\int_{-\phi_0}^{\phi_0} F_1(\beta)e^{jn\beta}d\beta. \quad (5)$$

The integral over ξ is just the space factor of a wire antenna carrying a "magnetic current" of linear density $G_1(z)$; except for the polarization it is the same as that of an electric current having the same distribution. The summation expresses the effect of the cylinder and the transverse distribution function.

Consider now the principal transverse plane, $\theta=\pi/2$. In this plane

$$E_\phi = \left[\frac{1}{2\pi^2}\frac{e^{-jkR}}{R}\int_{-l/2}^{l/2} G_1(\xi)d\xi\right]\sum_{n=-\infty}^{\infty}\frac{j^n e^{-jn\phi}}{H_n^{(2)\prime}(ka)} \times \int_{-\phi_0}^{\phi_0} F_1(\beta)e^{jn\beta}d\beta. \quad (6)$$

The bracketed term is constant with respect to ϕ; the radiation pattern in this plane is, therefore, given by the summation alone. The latter, except for constants which do not affect the pattern, is exactly the far-zone field (in the two-dimensional sense) produced by an infinite axial slot of width $2\phi_0$ in which the excitation is constant along the slot axis, directed transverse thereto, with a transverse distribution $F_1(\phi)$. Thus we have the general result that the pattern in the plane $\theta=\pi/2$ produced by a rectangular slot having only E_ϕ-excitation is independent of the axial distribution of E_ϕ and is the same

[4] C. J. Bouwkamp, Physica **XII**, 467 October, 1946.

[5] S. A. Schelkunoff, Phys. Rev. **56**, 308 (1939).

TABLE II. The function $\left| A \sum_{n=0}^{\infty} \frac{j^n \epsilon_n \cos n\phi}{\sin\theta H_n^{(2)\prime}(ka \sin\theta)} \right|$ for $ka=0.8$; $A=0.46667$ normalizes the function to unity at $\phi=0°$, $\theta=90°$.

θ \ ϕ	10°	20°	30°	40°	50°	60°	70°	80°	90°
0°	0.5942	0.6511	0.7506	0.8506	0.9248	0.9681	0.9876	0.9970	1.0000
10	0.5938	0.6493	0.7472	0.8460	0.9199	0.9634	0.9833	0.9930	0.9961
20	0.5924	0.6440	0.7369	0.8325	0.9052	0.9492	0.9702	0.9808	0.9841
30	0.5905	0.6355	0.7203	0.8103	0.8808	0.9252	0.9478	0.9596	0.9632
40	0.5880	0.6247	0.6984	0.7802	0.8469	0.8912	0.9154	0.9286	0.9327
50	0.5853	0.6123	0.6724	0.7434	0.8044	0.8473	0.8726	0.8870	0.8914
60	0.5826	0.5994	0.6441	0.7017	0.7544	0.7941	0.8195	0.8344	0.8390
70	0.5801	0.5871	0.6157	0.6574	0.6991	0.7333	0.7569	0.7713	0.7759
80	0.5782	0.5766	0.5895	0.6139	0.6421	0.6679	0.6875	0.7000	0.7041
90	0.5770	0.5686	0.5678	0.5751	0.5879	0.6027	0.6159	0.6248	0.6280
100	0.5766	0.5638	0.5526	0.5449	0.5423	0.5443	0.5488	0.5527	0.5543
110	0.5768	0.5623	0.5449	0.5262	0.5107	0.5004	0.4952	0.4931	0.4929
120	0.5777	0.5638	0.5445	0.5203	0.4965	0.4770	0.4635	0.4561	0.4542
130	0.5790	0.5675	0.5499	0.5255	0.4992	0.4758	0.4576	0.4472	0.4442
140	0.5806	0.5725	0.5591	0.5382	0.5142	0.4917	0.4731	0.4626	0.4594
150	0.5821	0.5778	0.5694	0.5540	0.5347	0.5161	0.4999	0.4912	0.4885
160	0.5835	0.5823	0.5787	0.5686	0.5544	0.5403	0.5272	0.5207	0.5188
170	0.5843	0.5854	0.5850	0.5787	0.5682	0.5574	0.5468	0.5419	0.5407
180	0.5846	0.5865	0.5872	0.5823	0.5731	0.5635	0.5539	0.5496	0.5486

TABLE III. The function $\left| A \sum_{n=0}^{\infty} \frac{j^n \epsilon_n \cos n\phi}{\sin\theta H_n^{(2)\prime}(ka \sin\theta)} \right|$ for $ka=2.5$; $A=0.13630$ normalizes the function to unity at $\phi=0°$, $\theta=90°$.

θ \ ϕ	10°	20°	30°	40°	50°	60°	70°	80°	90°
0°	0.7122	0.9167	0.9046	0.9722	0.9982	0.9852	0.9864	0.9952	1.0000
10	0.7087	0.9137	0.9027	0.9667	0.9948	0.9842	0.9846	0.9920	0.9962
20	0.6983	0.9043	0.8967	0.9505	0.9837	0.9807	0.9797	0.9835	0.9860
30	0.6814	0.8876	0.8889	0.9251	0.9624	0.9715	0.9717	0.9717	0.9722
40	0.6589	0.8626	0.8774	0.8931	0.9284	0.9518	0.9581	0.9571	0.9565
50	0.6320	0.8278	0.8621	0.8590	0.8819	0.9157	0.9325	0.9354	0.9355
60	0.6023	0.7823	0.8403	0.8271	0.8275	0.8603	0.8868	0.8970	0.8997
70	0.5720	0.7258	0.8080	0.7991	0.7750	0.7913	0.8179	0.8335	0.8389
80	0.5435	0.6595	0.7604	0.7710	0.7340	0.7239	0.7361	0.7487	0.7541
90	0.5194	0.5868	0.6943	0.7330	0.7042	0.6757	0.6655	0.6654	0.6669
100	0.5019	0.5138	0.6096	0.6741	0.6714	0.6466	0.6246	0.6120	0.6086
110	0.4922	0.4498	0.5113	0.5875	0.6151	0.6129	0.5989	0.5871	0.5828
120	0.4906	0.4060	0.4133	0.4763	0.5223	0.5453	0.5517	0.5503	0.5496
130	0.4956	0.3910	0.3389	0.3590	0.3984	0.4326	0.4539	0.4639	0.4673
140	0.5049	0.4032	0.3161	0.2776	0.2769	0.2954	0.3146	0.3269	0.3314
150	0.5155	0.4313	0.3456	0.2803	0.2341	0.2107	0.2022	0.2011	0.2012
160	0.5253	0.4616	0.3955	0.3420	0.2947	0.2602	0.2371	0.2234	0.2191
170	0.5319	0.4838	0.4358	0.4005	0.3688	0.3450	0.3298	0.3205	0.3179
180	0.5343	0.4918	0.4509	0.4230	0.3973	0.3801	0.3693	0.3627	0.3612

as that produced by an infinite axial slot having the same transverse distribution. This provides a rigorous basis for the conjecture of Sinclair,[6] Papas and King[7] and others that such a relationship exists in the case of a narrow rectangular slot in which $F_1(\phi)$ is a constant.

IV. THE NARROW RECTANGULAR SLOT

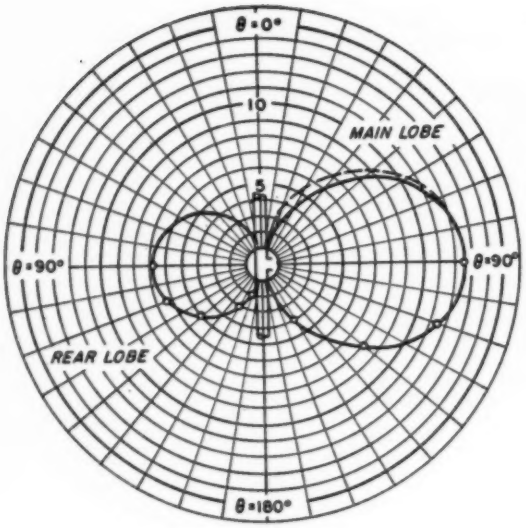
When the transverse dimension, $2a$, of the slot is small compared with the wave-length and the slot length, the significant mode of excitation is that in which there is only an E_ϕ-component that is uniform across the slot and has a sinusoidal distribution along its length. Such a slot is of particular interest because of its practical applications; the element a half-wave-length long is used extensively in microwave antennas.

It is customary to express the transverse characteristic $F_1(\phi)$ in terms of a voltage V; thus we have for the excitation

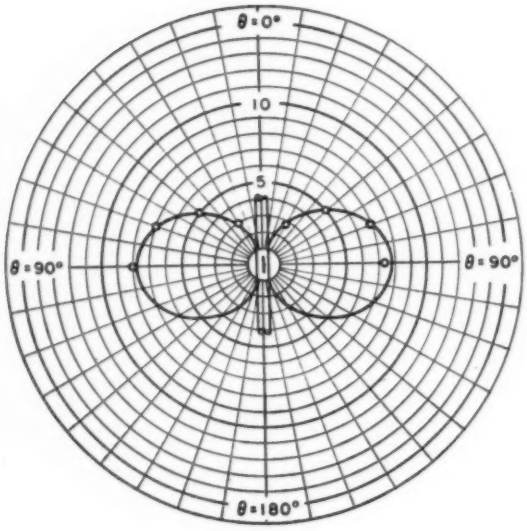
$$F_1(\phi)G_1(z)=(V/2a\phi_0)G_1(z). \tag{7}$$

On substituting into Eq. (5) we find the far-zone field to

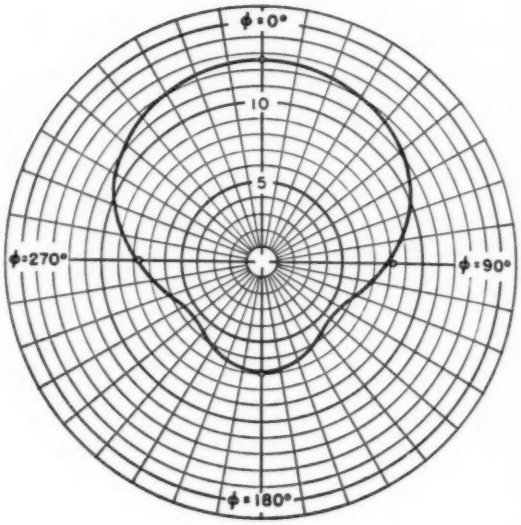
[6] G. Sinclair, Proc. I.R.E. **36**, 1487 (1948).
[7] C. H. Papas and R. W. P. King, Cruft Laboratory Technical Report No. 32, March 7, 1948.



(a)



(b)



(c)

FIG. 2. Radiation pattern ($|E|$) of the narrow half-wave slot; $ka=0.8$; (a) principal H-plane ($\phi=0, \pi$); (b) H-plane ($\phi=\pm\pi/2$); (c) principal E-plane ($\theta=\pi/2$). Solid line is the theoretical curve.

be

$$E_\phi=\frac{V}{2\pi^2 a}\frac{e^{-jkR}}{R}\int_{-l/2}^{l/2}G_1(\xi)\sin\theta e^{jk\xi\cos\theta}d\xi \times\sum_{n=-\infty}^{\infty}\frac{j^n e^{-jn\phi}}{\sin\theta H_n^{(2)\prime}(ka\sin\theta)}\frac{\sin n\phi_0}{n\phi_0}. \quad (8)$$

If ϕ_0 is sufficiently small, $\sin n\phi_0/n\phi_0\approx 1$ over the range of values of n for which the terms of the series are significant. This can be seen in Table I which shows the dependence on ka of the number of terms required for five figure accuracy and the associated value of ϕ_0 for which $\sin n\phi_0/n\phi_0\geq 0.9$ over the range of n. From a practical standpoint, therefore, we can ignore the $\sin n\phi_0/n\phi_0$ variation. Combining terms of positive and negative n, we can write finally for the radiation field of the narrow slot

$$E_\phi=\frac{V}{2\pi^2 a}\frac{e^{-jkR}}{R}\int_{-l/2}^{l/2}G_1(\xi)\sin\theta e^{jk\xi\cos\theta}d\xi \times\sum_{n=0}^{\infty}\frac{j^n\epsilon_n\cos n\phi}{\sin\theta H_n^{(2)\prime}(ka\sin\theta)};\quad \begin{cases}\epsilon_n=1, & n=0,\\ =2, & n>0.\end{cases} \quad (9)$$

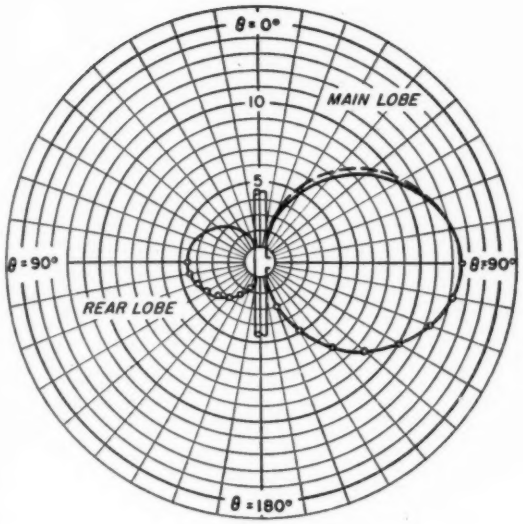
In its reduced form the series is a space factor dependent on the cylinder alone. With the series tabulated as a function of ka, θ, and ϕ the computation of the field produced by narrow axial slots becomes a simple straightforward task. Tables II and III give the function in a $10°\times 10°$ mesh in θ and ϕ for two values of ka used in our experimental work. The latter correspond to values that are likely to be used for microwave antennas.

The Half-Wave-Length Slot

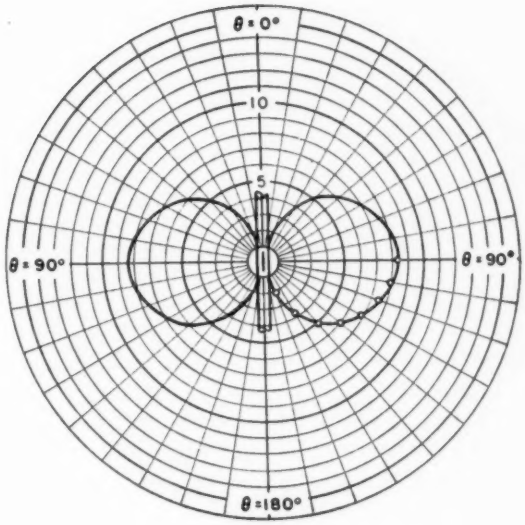
When the length of the slot is $\lambda/2$, $G_1(z)=\cos(\pi z/l)$ and the far-zone field is

$$E_\phi=\frac{V}{2\pi^2 a}\frac{e^{-jkR}}{R}\left[\frac{\cos(\pi/2\cos\theta)}{\sin\theta}\right]\times\sum_{n=0}^{\infty}\frac{j^n\epsilon_n\cos n\phi}{\sin\theta H_n^{(2)\prime}(ka\sin\theta)}. \quad (10)$$

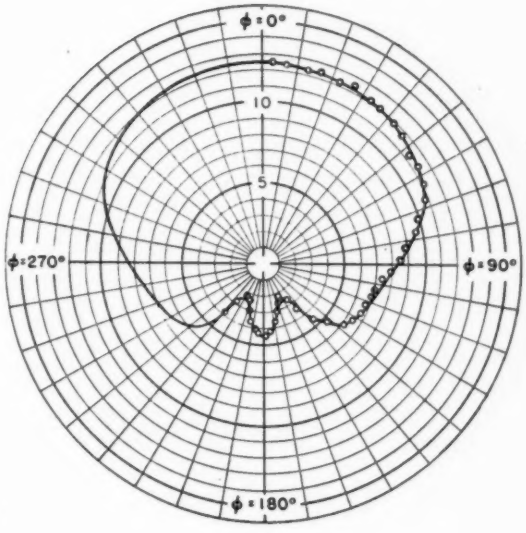
The bracketed term will be recognized to be the space factor of a half-wave-length wire antenna. The effect of the cylinder can be seen in the patterns shown in Figs. 2 and 3. Figures (2a) and (3a) are principal H-plane patterns ($\phi=0, \pi$); the dashed line is one quadrant of the pattern of the half-wave radiator in free-space. It is seen that the main lobe of the slot pattern is closer to the half-wave radiator pattern in the case of the larger cylinder than in the case of the smaller one. The corre-



(a)



(b)



(c)

FIG. 3. Radiation pattern ($|E|$) of the narrow half-wave slot; $ka=2.5$; (a) principal H-plane ($\phi=0, \pi$); (b) H-plane ($\phi=\pm\pi/2$); (c) principal E-plane ($\theta=\pi/2$).

spondence may be expected to increase with increasing cylinder radius, for the pattern of a corresponding slot in an infinite plane sheet is just that of the half-wave radiator. The principal transverse patterns ($\theta=\pi/2$) are given in Figs. (2c) and (3c).

The circles are the experimental points. The excellent agreement between the measured and theoretical values shown here was also obtained in the many other patterns that were investigated.[8] Figure 2 pertains to a probe-fed slot of width 0.006 in. cut in the wall of a $\frac{5}{16}$ in. O.D. coaxial line. The operating wave-length of 3.12 cm was chosen so that the value of ka would coincide with one of those for which transverse plane patterns were measured by Sinclair[6] in his study of the relation between the finite slot and the infinite slot. The computed pattern shown in Fig. (2c) will be found to be in excellent agreement with Sinclair's data and the pattern of an infinite slot.

The patterns in Fig. 3 were obtained for a slot of width .012 in. cut into the wall of a circular wave guide of O.D. 1 in. The circular wave guide was fed by way of a standard type of transition section from a rectangular wave guide and the slot was cut along the extended center line of the narrow side of the rectangular guide. The operating wave-length was 3.2 cm.

[8] The reader is referred to our report, "The radiation from a rectangular slot in a circular cylinder," Univ. of Calif. Antenna Laboratory Report No. 158. (Bu Ships Contract NO-bsr-39401) October 26, 1949 for a more complete set of patterns.

V. ACKNOWLEDGMENTS

We take pleasure in expressing our appreciation to Mr. J. Priedigkeit of the Antenna Laboratory for his assistance with the measurements. The work was supported in part by the Bureau of Ships under Contract No-bsr-39401 with the University.

The Thermal Conductivity of Soils

ANDREW GEMANT
Research Department, The Detroit Edison Company, Detroit, Michigan
(Received January 31, 1950)

The thermal conductivity of soils of varying moisture content is calculated on a physical basis. The results of this computation are, in a range of moisture contents from 5 to 25 percent, in agreement with experimental data in the literature. The influence of soil composition on conductivity is also explained by the theory. The use for practical purposes of an average resistivity value of 60 to 70 thermal ohms appears justified in the light of this analysis.

THE thermal conductivity of soils is an essential factor in determining the heat dissipation in various industrial installations in the ground. Such installations are, among others, underground high voltage cable lines, steam lines, and cross-country oil pipe lines.

For the design and operation of such installations a knowledge of the conductivity of the soil is necessary. The value varies with the composition of the soil, whether sand, loam, or clay, but primarily it depends on the moisture content and thus varies with the season. Pertinent data referring to various soils are presented by Nelson.[1]

For many purposes, particularly in the cable field, it is customary to use average values. An often used value for the resistivity has been 120 thermal ohms, whereas in recent years a value of 80 is more frequently used. Recent determinations by Sanderson, Sticher, and McGrath[2] indicate that still lower values, about 60 thermal ohms, might be justified.

Because of the importance of soil conductivity, it was thought of interest to calculate its value on a physical basis. Such a computation might also indicate the right order of magnitude of the average figures to be used for practical purposes.

For the following considerations both thermal conductivities λ and resistivities ρ will occur. The former is usually expressed in cal./cm sec. °C, the latter in cm °C/watt, or thermal ohms. In these units

$$\rho=0.239/\lambda. \tag{1}$$

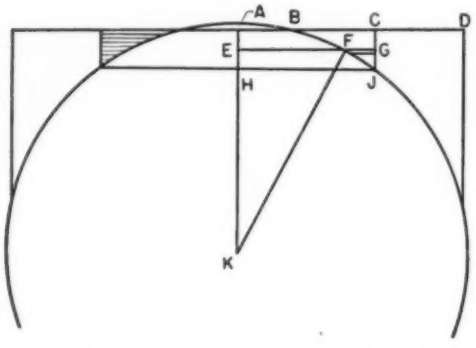


FIG. 1. Assumed geometric arrangement of solid grains and water in soil.

As mentioned, λ increases with increasing moisture content. For dry soil, λ is in the order 0.0007, increasing with increasing moisture content up to about 0.005. An apparent paradox may be mentioned at this point. One would expect that, since water increases the thermal conductivity, water itself should have a conductivity higher than soaked soil. This, however, is not the case, λ for water being 0.0014.

This apparent contradiction can be explained in the following manner. Dry soil has a low conductivity because air, a poor thermal conductor, separates the solid grains of the soil. If one adds water to the soil, hence gradually replacing the air by water, the conductivity increases. Finally, we arrive at soil saturated with water. Of the two components of such a system, the solid has a higher conductivity than the liquid; hence replacing the solid by the liquid will result in lower conductivity. The statement above, namely, that the solid component of soils is a better thermal conductor than water, is corroborated by data, as will be shown below.

The calculation of the conductivity of composite structures[3,4] can be carried out with a greater or smaller degree of rigorousness. Absolute rigorousness is impossible because of the irregularity of the shape and distribution of the grains. For the present purpose an approximate calculation was deemed sufficient.

The soil is assumed as being composed of solid spherical particles of radius $KF=r$ (Fig. 1). In reality, the grains are more likely to be oblate ellipsoids. The arrangement of the spheres is assumed such that each sphere touches six neighbors. In reality, the packing will be closer; our arrangement, however, does not deviate too much from reality, since it yields a porosity of 43 percent against about 38, found experimentally. If the spheres were to touch at a geometrical point only, the resistivity of dry soil would be close to that of air which is contrary to experience. The spheres, therefore, are assumed to be chipped and contact over the circular areas of radius $AB=a$. For the sake of simplicity the length $AK=AD$ is termed unity; the cubical volume housing the chipped sphere is then 8. The actual grain size is not involved in the following calculations.

[1] W. L. Nelson, Oil and Gas J. **44**, 110 (September, 1945).
[2] Sanderson, Sticher, and McGrath, Trans. A.I.E.E. **67**, 487 (1948).
[3] J. B. Austin, *Symposium on Thermal Insulating Materials* (American Society for Testing Materials, Philadelphia, 1939).
[4] W. Meissner and G. U. Schubert, "Physics of solids I," FIAT Rev. German Sci., 212 (1947).

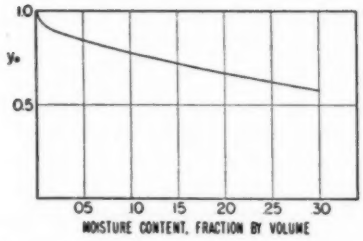


FIG. 2. Thickness of water layer between grains as a function of moisture content.

If moisture is present, it will, owing to capillarity,[5] be located around the circles of radius a, occupying a wedge-shaped ring, the cross section of which is BCJ, shown also by the cross-hatched area in Fig. 1. The volume of the ring increases with increasing moisture content. If h, the moisture content, designates the ratio of water volume to total volume, then the volume of the ring BCJ equals $8h/6=1.33h$.

Any point F on the sphere is characterized by its coordinates $EF=x$ and $EK=y$. The water reaches down to point J, characterized by its coordinates $HJ=x_0$ and $KH=y_0$. Obviously, y_0 is a function of the moisture content h, and for the subsequent derivation this correlation is needed. Its calculation is straightforward, as shown.

The volume of an infinitesimally thin ring-shaped layer (Fig. 1), the cross section of which is shown at FG, is given by $\pi(x_0^2-x^2)dy$, or, because of

$$x^2+y^2=r^2, \quad \pi(y^2-y_0^2)dy.$$

The volume of the total ring is given by

$$\pi\int_{y_0}^{1}(y^2-y_0^2)dy.$$

Solving the integral, one obtains:

$$1.33h=\pi(0.33-y_0^2+0.67y_0^3). \tag{2}$$

For moisture contents over about 0.2 the formula is not strictly correct, since there will be an overlapping of the neighboring ring-shaped water pools.

Figure 2 is a graphical presentation of Eq. (2) for moisture contents up to 0.30. It may be noted that $(1-y_0)$ is a measure of the thickness of the water layer between grains. Use will be made later of this graph.

The thermal resistivity is computed by taking AK as the direction of heat flow and integrating the resistance of infinitesimally thin layers normal to the y axis. Since the flow lines are not parallel to y but diverge below y_0, this method is approximate.

Consider again the elementary layer at F. If the conductivity of the solid phase is λ_s and that of the water λ_w, the resistance of the inner circle is

$$dy/(\pi x^2\lambda_s)$$

and that of the outer ring

$$dy/[\pi(x_0^2-x^2)\lambda_w].$$

The resistance of both together is obtained by adding the conductances and taking the reciprocal of the sum. By replacing x by y, and considering that $r^2=1+a^2$, one obtains

$$dy/[\pi(f^2-g^2y^2)],$$

where

$$f^2=(1+a^2)\lambda_s-y_0^2\lambda_w \tag{3}$$

and

$$g^2=\lambda_s-\lambda_w. \tag{4}$$

The resistance of all layers between $y=1$ and y_0 is then:

$$\frac{1}{\pi}\int_{y_0}^{1}\frac{dy}{f^2-g^2y^2}=\frac{1}{2\pi fg}\log\frac{(f+g)(f-gy_0)}{(f-g)(f+gy_0)}. \tag{5}$$

Disregarding chipping at four points of the "equator" of our sphere, the resistance of all layers between $y=y_0$ and 0 is obtained as

$$\frac{1}{2\pi f'g'}\log\frac{f'+g'y_0}{f'-g'y_0}, \tag{6}$$

where

$$f'^2=(1+a^2)\lambda_s \tag{7}$$

and

$$g'^2=\lambda_s. \tag{8}$$

The resistance of all layers between $y=0$ and 1 is given by the sum of (5) and (6). Since the cross-sectional area of this volume is 4, the resistivity of the structure is four times greater. Thus

$$\frac{1}{\lambda}=\frac{2}{\pi fg}\log\frac{(f+g)(f-gy_0)}{(f-g)(f+gy_0)}+\frac{2}{\pi f'g'}\log\frac{f'+g'y_0}{f'-g'y_0}. \tag{9}$$

In the derivation above, the conductance of the air phase was neglected, as well as the conductance of the four water rings located at four points of the "equator" of our sphere.

For the numerical calculations which follow, numerical values of λ_s, λ_w, and a are required. The value of

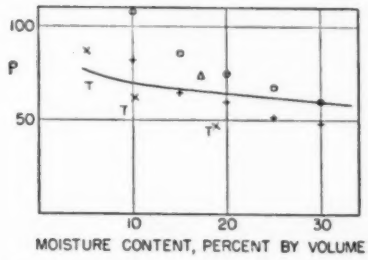


FIG. 3. Resistivity of soil *vs.* moisture content. (Solid line is calculated, points are experimental data from literature.)

[5] M. Devienne, J. Rech. Centr. Nat. Rech. Sci., No. 8, 238 (1949).

TABLE I. Influence of composition on conductivity of soils. Density = 1.75; moisture content = 17.5 percent.

Soil	ρ observed ohms	λ_s cal./cm sec. °C	ρ calculated ohms
Sand	46	0.014	50
Sandy loam	69	0.010	66
Silt loam	99	0.0058	101

y_0 can be read from Fig. 2 if the moisture content of the soil is known.

Although the chief constituent of soils is silica, the conductivity of pure quartz crystals is not an appropriate choice for λ_s, since the grains of soils are not composed of pure quartz crystallites. Closer to reality is probably quartzite, a compact rock composed of quartz; still better perhaps, quartz-schist, a metamorphosed rock essentially of quartz; or sandstone, a rock essentially of sand and some natural cement. The conductivity of quartzite[6] is 0.014, that of quartz-schist[7] 0.0097, and that of sandstone[7] 0.011. As an average of the last two data, $\lambda_s = 0.010$ was taken. The value of λ_w, as mentioned earlier, is 0.0014.

Concerning the radius a of the circle of contact, its value can be computed from the conductivity of dry soil, λ_d, a representative value of which is 0.00066. For this case, $1/\lambda_d$ is given by the second term of Eq. (9) with $y_0 = 1$. If, now, $a \ll 1$, the corresponding equation reduces to

$$\frac{1}{\lambda_d} = \frac{2}{\pi \lambda_s} \log \frac{4}{a^2}, \tag{10}$$

from which, with $\lambda_s = 0.010$, one obtains $a = 2 \times 10^{-5}$. The area of contact is very small compared with the grain dimension, and may be neglected in the numerical computation; in other words, unity can be substituted for $(1+a^2)$ in Eqs. (3) and (7).

Equation (9) now can be used for numerical computation of the conductivity λ of soils of varying moisture content. The result, in a range of $h = 0.05$ to 0.3, that is 5 to 30 percent by volume of moisture, is shown by the curve of Fig. 3. The ordinates are ρ in thermal ohms, calculated from λ by Eq. (1).

For the sake of comparison with the calculated curve, experimental data of Krischer[6] on two samples of soil are plotted on Fig. 3. The soils measured were loamy sand of average grain size 0.085 cm. The crosses in Fig. 3 refer to a sample of density (dry) 1.6 and 38 percent porosity, the circles to a sample of density 1.7 and 34 percent porosity. In addition, a value indicated by a triangle obtained by Sanderson *et al.*[2] is plotted.

[6] O. Krischer, Beihefte z. Gesundheits—Ing. Series I, No. 33 (1934).
[7] H. A. Nancarrow, Proc. Phys. Soc. 45, 447 (1933).

A few pertinent data taken from a recent paper by M. S. Kersten[8] are also plotted on Fig. 3. They refer to a sandy soil of density 1.7 and 37 percent porosity (X) and of density 1.8 and 34 percent porosity (T).

Above five percent moisture content, the agreement between experimental and calculated data is good; below five percent, experimental data are higher. The calculation involves rather specific assumptions as to the geometry of contact points between grains; for low moisture contents where the contacts become increasingly significant, the calculation becomes less reliable. The moisture range of greatest practical importance is between 5 and 25 percent, and for this range the calculation corroborates the experimental figures rather well. If the present calculation were carried out for oblate ellipsoids, instead of spheres, or for grains bounded partly by planes, the agreement in the low moisture content range might perhaps be improved.

The influence of the composition of the soil on conductivity, according to Kersten's observations, is such that λ increases in the sequence clay—loam—sand. Soils of constant density and moisture content must be chosen for such a comparison. Table I gives in the first column the designation of the soil and in column 2 the resistivity in ohms, as found by Kersten.[8]

According to the theory given, the influence of the composition is represented by λ_s which for a sandy loam, as discussed, is 0.010. Sand is essentially quartz for which the conductivity of quartzite, 0.014, is appropriate. Silt loam consists to a large extent of kaolin which is essentially finely divided feldspar, aluminum silicate. For feldspar,[9] $\lambda_s = 0.0058$. These values are listed in the third column of Table I. The resistivities, as calculated from Eq. (9), are given in column 4; the agreement with the observed data is satisfactory.

In practical instances the conductivity of the sequence clay—loam—sand may be reversed, clay having the highest conductivity. This is not in contradiction with the data given in Table I, which refers to the same moisture content. The reversal is caused by the varying ability of soils in retaining moisture. Thus sand, having large grains, passes water easily and has, as a rule, a low moisture content; hence, a resistivity of 80 ohms or more will result. Clay, on the other hand, has very fine pores, water is easily retained and the moisture content, as a rule, is high; the resistivity may be as low as 60 or even 50 ohms. The average soil, being a sandy loam of a water content of about 20 percent, has, according to Fig. 3, a resistivity of 60 to 70 thermal ohms.

[8] M. S. Kersten, University of Minnesota Eng. Exp. Station Bulletin No. 28 (1949).
[9] Landolt-Börnstein, *Physik-Chem. Tabellen* (Verlag. Julius Springer, Berlin, Edwards' Lithoprint, 1943), Vol. II, p. 1295.

Theory of Blind Navigation by Dynamical Measurements

J. J. Gilvarry, S. H. Browne,* and I. K. Williams
The Rand Corporation, Santa Monica, California
(Received February 1, 1950)

The differential equation is considered which determines the position of a vehicle from dynamical measurements of the non-gravitational acceleration **b** made internally. Three linear approximations to the gravitational field **g**(**r**) of the earth, which lead to explicit solutions of this equation, are considered and their limitations are discussed. An interval-wise solution (linear continuation) for trajectories of extended range is described, which is based on such linear approximations and has definite advantages in this application. The theory is applied to the trajectory of the German A10 vehicle.

CURRENT developments in rocket and jet propulsion have made blind navigation a matter of practical importance. This paper discusses the theory of blind navigation by means of dynamical measurements (measurements of forces or accelerations) made on a proof body in a reference frame internal to a vehicle. Instrumentally, such measurements are made by an accelerometer in a reference frame provided by gyroscopic or other means. A navigation system essentially of the type considered was used in the German V2. In the following discussion, the vehicle will be idealized as a windowless box.

Under the assumption of local uniformity of the gravitational field, the information derivable from the accelerometer is limited[1] by the applicability of Einstein's equivalence principle. In fact, this limitation was recognized by the investigator[2] who built the first (consciously-designed) accelerometer.

I. THE g-CORRECTION EQUATION

The accelerometer reference frame Q will be taken non-rotating for convenience, with origin at a point Q in the box. If $\mathbf{R}$ is the radius vector of the point Q referred to an inertial reference frame O, then the acceleration $d^2\mathbf{R}/dt^2$ of this point must consist of two parts: a part $\mathbf{g}$ due to the local gravitational acceleration in space, and a part $\mathbf{b}$ due to the operation of non-gravitational forces. Thus one has

$$d^2\mathbf{R}/dt^2=\mathbf{g}+\mathbf{b}. \tag{1}$$

If $\boldsymbol{\rho}$ is the radius vector in the frame Q of the proof body and m is its mass, the equation of motion of this body in the inertial frame O is

$$d^2(\mathbf{R}+\boldsymbol{\rho})/dt^2=\mathbf{g}+\mathbf{L}/m, \tag{2}$$

where $\mathbf{L}$ is the force on the proof body linking it to the frame Q. Subtraction of (1) from (2) yields

$$d^2\boldsymbol{\rho}/dt^2-\mathbf{L}/m=-\mathbf{b} \tag{3}$$

for the equation of motion of the proof body in the observer's frame Q. If the quantity on the left-hand side in Eq. (3) is defined as the reading of the accelerometer, it follows that an accelerometer reads the negative of the acceleration of the origin of its associated reference frame due to non-gravitational forces. From the method of derivation, it is clear that Eq. (3) holds independently of variation of mass, rotation, or non-rigidity of the box.

In Eq. (1) only the term $\mathbf{b}$ is susceptible to physical measurement in the box, and the observer can integrate the differential equation to yield his position vector $\mathbf{R}(t)$ only if he knows the functional form of $\mathbf{g}$ in its dependence on position coordinates and also the initial position and velocity of the box. When the term $\mathbf{g}$ is replaced by the function specifying its dependence on position coordinates in space, and the term $\mathbf{b}$ is considered a function $\mathbf{b}(t)$ of time, Eq. (1) will be referred to as the g-correction equation.

For motion of the box in the earth's neighborhood, the inertial frame O can be taken as approximately geocentric. It is convenient to select an initial point O_0 with radius vector $\mathbf{R}_0$ which, in general, is outside the earth's surface and on the curve of motion C (Fig. 1). The gravitational field corresponding to a spherical earth is

$$\mathbf{g}=-g_0(R_0^2/R^3)\mathbf{R}, \tag{4}$$

if g_0 is the value of g at O_0. The g-correction equation then becomes

$$d^2\mathbf{R}/dt^2+g_0(R_0^2/R^3)\mathbf{R}=\mathbf{b}(t). \tag{5}$$

The sphere S_0 (Fig. 1) through O_0 and concentric with the earth will be referred to as the reference sphere. For an initial point on the actual (approximately spheroidal) earth, the radius and surface gravitation of the reference sphere can be taken as the mean radius R_e and mean gravitational acceleration g_e over the earth's surface, which are defined[3] in first approximation by

$$R_e=(2/3)R_{eq}+(1/3)R_{pol}, \tag{6a}$$

$$g_e=(2/3)g_{eq}+(1/3)g_{pol}, \tag{6b}$$

where R_{eq} and R_{pol} are the equatorial and polar radii

* Now at North American Aviation, Inc., Los Angeles, California.

[1] J. J. Gilvarry, Phys. Rev. **73**, 1409 (1948). A misprint exists in this letter; in the first line of the second column of p. 1410, delete "vertically."

[2] F. W. Lanchester, Proc. Phys. Soc. **19**, 691 (1905).

[3] G. G. Stokes, Camb. and Dubl. Math. J. **4**, 194 (1849); Trans. Camb. Phil. Soc. **8**, 672 (1849); also *Mathematical and Physical Papers* (Cambridge University Press, Cambridge, 1883), Vol. II, p. 104.

respectively of the earth, and g_{eq} and g_{pol} are the corresponding gravitational accelerations. The radius R_e is, in first approximation, equal to the radius of the earth's sphere of equal volume, and defines an equivalent sphere S_e (Fig. 1) with surface gravity g_e for the earth.[4] Equation (5) is valid in a geocentric frame O' fixed in the rotating earth if the centrifugal and Coriolis accelerations (and the latitude-dependent component of $\mathbf{g}$) can be neglected.[5]

Let a Cartesian frame O_0 (Fig. 1) be defined with origin at the initial point of motion, such that the z-axis is vertical, the x-axis is horizontal in the plane determined by the z-axis and the initial velocity vector $\mathbf{v}_0$, and the y-axis is selected to make the frame right-handed. It will be assumed for simplicity in what follows that the initial velocity $\mathbf{v}_0$ is non-vanishing unless otherwise stated (and hence the convention above defines the direction of the x-axis except when the initial velocity is in the vertical direction). If $\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$ is the radius vector of the box in the frame O_0, Eq. (5) becomes

$$d^2\mathbf{r}/dt^2+\omega_0^2(R_0/R)^3(R_0\mathbf{k}+\mathbf{r})=\mathbf{b}(t), \tag{7}$$

where the parameter ω_0 is defined by

$$\omega_0=(g_0/R_0)^{\frac{1}{2}}, \tag{8}$$

and

$$R=[x^2+y^2+(z+R_0)^2]^{\frac{1}{2}}, \tag{9}$$

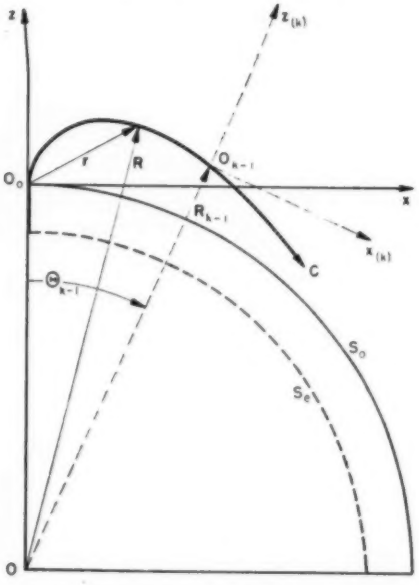


FIG. 1. Coordinate frames.

[4] The equality of volumes follows on the assumption that squares and higher powers of the ellipticity $e=(R_{eq}-R_{pol})/R_{eq}$ are negligible compared to unity. If the density of the earth's crust is assumed constant, R_e is likewise the radius of the earth's sphere of equivalent mass. Hence, Eq. (4) with $R_0=R_e$ and $g_0=g_e$ yields the correct asymptotic value GM/R^2 (G=gravitational constant, M=mass of the earth) of g at large distances from the earth.

[5] Equation (5) likewise ignores a slight perturbation due to lunar and solar attraction (similar in origin to the perturbations producing tides). The corresponding effect for artillery shells is discussed in F. R. Moulton, *New Methods in Exterior Ballistics* (University of Chicago Press, Chicago, 1926), p. 19.

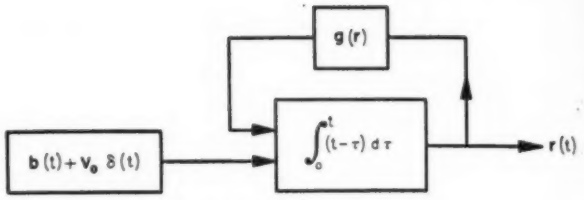


FIG. 2. Computer diagram.

with $\mathbf{b}=b_x\mathbf{i}+b_y\mathbf{j}+b_z\mathbf{k}$. The initial conditions are $\mathbf{r}_0=0$ and $\mathbf{v}_0=v_{x,0}\mathbf{i}+v_{z,0}\mathbf{k}$.

Equation (7) can be formulated as an integral equation

$$\mathbf{r}=\mathbf{v}_0t+\int_0^t(t-\tau)[\mathbf{b}(\tau)+\mathbf{g}(\mathbf{r})]d\tau, \tag{10}$$

in which $\mathbf{g}(\mathbf{r})$ is $\mathbf{g}$ of (4) with its dependence on components of $\mathbf{r}$ represented symbolically, and the cubature operator $\int_0^t(t-\tau)d\tau$ performs the operation of double integration with respect to time (as integration by parts verifies). From the standpoint of an automatic computer in the box, the sequence of operations shown diagrammatically in Fig. 2 yields the position vector $\mathbf{r}(t)$ from $\mathbf{b}(t)$ (the initial velocity $\mathbf{v}_0$ has been put in symbolically by means of the Dirac δ-function, $\delta(t)$, which is singular at $t=0$).

From the standpoint of determining $\mathbf{r}(t)$ as a closed solution in the time, Eq. (7) poses serious difficulties since $\mathbf{b}(t)$ is an arbitrary function of time and the equation is non-linear in $\mathbf{r}$. In the general case, integrals of energy and angular momentum cannot be found except as purely formal expressions. Hence, the classical method of using integrals of the motion to reduce the order of the component differential equations is not possible in general. Exact solutions in some particular cases are available; if $\mathbf{b}$ is zero, for example, Eq. (7) reduces to the differential equation of an elliptical free-flight trajectory. In practical cases, however, $\mathbf{b}$ and $\mathbf{g}$ are of comparable magnitude over the trajectory as a whole (see Fig. 4), so that perturbation methods are not generally useful. Resort can be made to various numerical and iterative methods. A difference method due to Hartree,[6] which avoids intermediate computation of the velocities, can be mentioned. The integral equation, (10), can be identified in coordinate formulation with a system of non-linear Volterra integral equations of the second kind[7] to yield a general solution by iteration. Although iterative solutions have obvious limitations, a solution of this type has been applied[8] as a computational method for specific $\mathbf{b}(t)$.

In the practical problem of computer design, the non-linear $\mathbf{g}$-term of Eq. (7) presents design complications. Furthermore, a suitable approximation for $\mathbf{g}$ may be sufficient on trajectories of restricted range, or when

[6] D. R. Hartree, Mem. Manch. Lit. Phil. Soc. **77**, 91 (1932).

[7] V. Volterra, *Leçons sur les Equations Intégrales* (Gauthier-Villars, Paris, 1913), p. 90.

[8] E. T. Benedikt, Phys. Rev. **74**, 1213(A) (1948).

the motion is constrained to a particular surface (see Section 2). The simplest such approximation consists in replacing $\mathbf{g}$ by its constant value $-g_0\mathbf{k}$ at O_0, which yields the zero-order solution

$$\mathbf{r}=\mathbf{v}_0t-(g_0t^2/2)\mathbf{k}+\int_0^t(t-\tau)\mathbf{b}(\tau)d\tau, \qquad (11)$$

valid in a restricted neighborhood of the origin. The remainder of this paper will consider the problem of replacing $\mathbf{g}$ by linear approximations. These linear approximations will be made the basis of an interval-wise solution having certain advantages.

II. THE SPHERICAL SOLUTION

The g-correction equation, (7), has a simple and exact solution when the vehicle is constrained to move on the surface of a sphere concentric with the earth, and the corresponding solution will be referred to as a spherical solution. The sphere of motion will be taken as the reference sphere of radius R_0, on which the outward-drawn unit normal is $\mathbf{R}/R_0$ or $\mathbf{k}+\mathbf{r}/R_0$. The condition on $\mathbf{b}$ for motion on S_0 is

$$\mathbf{b}\cdot(R_0\mathbf{k}+\mathbf{r})=g_0R_0^2/R-v^2, \qquad (12)$$

which is necessary and sufficient if the initial velocity vector is tangent to S_0 at $O_0(\mathbf{v}_0=v_0\mathbf{i})$. Physically, Eq. (12) states (when divided by R) that the radial component of $\mathbf{b}$ equals the gravitational acceleration less the centrifugal acceleration on a sphere of radius R. In practice, the condition (12) can be met by means of an altimeter (of atmospheric-pressure or radar type) used to control independently the altitude of the vehicle relative to the earth. The field $\mathbf{g}_s$ on the surface of S_0 is

$$\mathbf{g}_s=-g_0\mathbf{k}-\omega_0^2\mathbf{r}, \qquad (13)$$

which will be referred to as the spherical field.

The g-correction equation in the spherical case is

$$d^2\mathbf{r}/dt^2+\omega_0^2\mathbf{r}=\mathbf{b}-g_0\mathbf{k}. \qquad (14)$$

The solution of Eq. (14) subject to the condition (12) on $\mathbf{b}$ is

$$\mathbf{r}=[(\sin\omega_0t)/\omega_0]\mathbf{v}_0-[2R_0\sin^2\omega_0t/2]\mathbf{k}+(1/\omega_0)\int_0^t\mathbf{b}(\tau)\sin\omega_0(t-\tau)d\tau, \qquad (15)$$

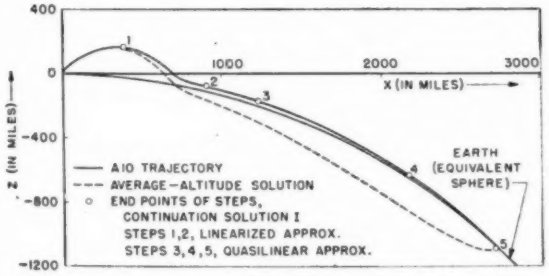


FIG. 3. Trajectory of A10 vehicle.

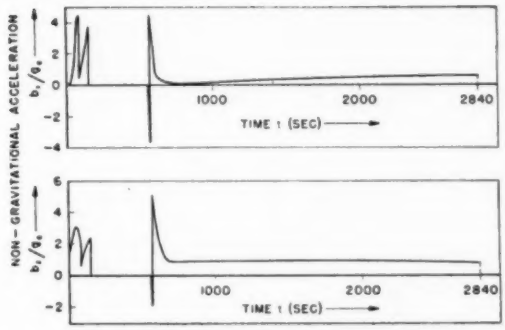


FIG. 4. Non-gravitational accelerations b_x and b_z on A10 vehicle in terms of g_e (surface gravity on equivalent sphere).

with the restriction that $\mathbf{v}_0=v_0\mathbf{i}$. If the sine functions appearing in the spherical solution are replaced by the leading terms of their Taylor expansions, the spherical solution (15) reduces to the zero-order solution (11).

As a trivial example of the spherical solution, consider a box moving on the surface of the reference sphere S_0 and such that $\mathbf{b}=0$. From Eq. (12), the speed $v_{s,0}$ is a constant

$$v_{s,0}=(g_0R_0)^{\frac{1}{2}}. \qquad (16)$$

Equation (15) in this case corresponds to a circle of radius R_0 described with the constant angular velocity ω_0 and the period

$$T_0=2\pi(R_0/g_0)^{\frac{1}{2}}, \qquad (17)$$

which is approximately 84.4 min. for an origin on the equivalent sphere S_e. Hence, this special case corresponds physically to the motion of a satellite around the earth in a circular orbit, and the speed $v_{s,0}$ will be referred to as the satellite (or circular) speed corresponding to the radius R_0.

When the constraining condition (12) is not satisfied exactly at all times, the altitude,

$$h=R-R_0, \qquad (18)$$

of the vehicle relative to the reference sphere S_0 is non-vanishing, and the spherical solution (15) yields a corresponding error. If $\delta_s\mathbf{r}$ is the true solution of Eq. (7) less the spherical solution of (15) for a given $\mathbf{b}$, then one has

$$d^2\delta_s\mathbf{r}/dt^2+\omega_0^2\delta_s\mathbf{r}=g_0\xi(\mathbf{k}+\mathbf{r}/R_0), \qquad (19)$$

where the error parameter ξ is defined by

$$\xi=1-(R_0/R)^3\sim 3h/R_0, \qquad (20)$$

and the approximation indicated is valid for $h\ll R_0$. The quantity $\xi(\mathbf{k}+\mathbf{r}/R_0)$ is the gradient of a harmonic function, and hence its magnitude assumes a maximum value on the boundary of any closed region of space (not including the center of the earth) considered. Thus, if a uniform bound H on the absolute value of h is known, then

$$M=(1-H/R_0)^{-2}-(1-H/R_0)\sim 3H/R_0 \qquad (21)$$

is a uniform bound on $|\xi(\mathbf{k}+\mathbf{r}/R_0)|$ for $H<R_0$, and hence one can show that

$$|\delta_s\mathbf{r}| \leqslant 2g_0M \sin^2\omega_0 t/2, \quad (t \leqslant T_0/2), \tag{22}$$

if no error exists in the knowledge of initial conditions (the direction of $\mathbf{v}_0$ is now unrestricted). Since $|\delta_s\mathbf{r}|$ vanishes as $M\to 0$ (or $H\to 0$), the spherical solution (15) is a valid approximation in the large (or "on the average") for motion which is closely spherical, within the limitations of (22) or a closer error bound.

Over a trajectory of extended range, the spherical solution is not useful if significant systematic variations of altitude exist. The difficulty usually cannot be avoided by taking an average R corresponding to an average altitude in the $\mathbf{g}$-term of Eq. (7); this point can be illustrated by an example. Figure 3 shows the trajectory (solid curve) of the German A10 vehicle,[9,10] which is extreme in its altitude variation (185 mi.), and Fig. 4 shows the non-gravitational accelerations[11] b_x and b_z for this vehicle as a function of time. For this trajectory, a solution of the g-correction equation is shown in Fig. 3 (dashed curve) which corresponds to an average altitude of 16.9 mi. (i.e., R was taken as $R_e+16.9$ mi. in the $\mathbf{g}$-term of Eq. (7)). This particular average-altitude solution yields a negligible error at the terminal point of the motion but defines a trajectory passing under the earth's surface.

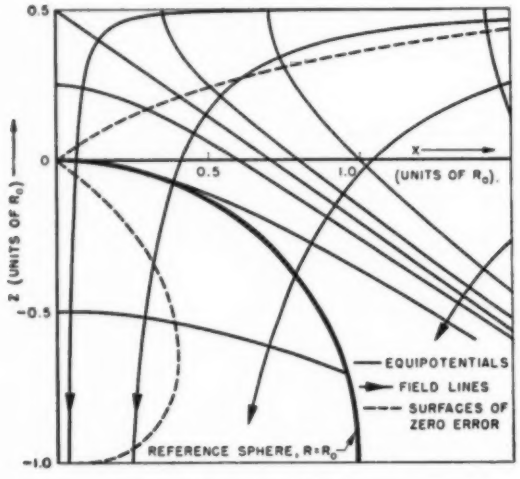


FIG. 5. Linearized field diagram.

[9] W. G. A. Perring, J. Roy. Aero. Soc. **50**, 483 (1946).

[10] J. M. J. Kooy and J. W. H. Uytenbogaart, *Ballistics of the Future* (McGraw-Hill Book Co., Inc., New York, 1946), p. 399.

[11] Only a limited amount of information is available on the A10 (this vehicle was proposed but never built, and was to consist essentially of a winged V2 equipped with a booster rocket). The trajectory of Fig. 3 fits the A10 as closely as the available information permits. The quantities b_x and b_z were determined by differentiation from the assumed trajectory. Of the peaks in each curve of Fig. 4, the first two correspond to release of the booster and power cut-off in that order, and the ones at 560 seconds correspond to re-entry into appreciable atmosphere. Rotation of the earth and latitude variation of g have been completely neglected in the data of Figs. 3 and 4.

III. THE LINEARIZED APPROXIMATION

It is clear from the preceding discussion that the essential problem in connection with the g-correction equation exists when significant variations in altitude above the earth occur. This section discusses an approximate analytic solution which is superior to the zero-order approximation (11) when variations in altitude above the earth occur.

The field $\mathbf{g}(\mathbf{r})$ of (4) can be expanded about the origin O_0 to yield

$$\mathbf{g}(\mathbf{r}) = -g_0\mathbf{k} - \omega_0^2[\mathbf{r} - 3(\mathbf{k}\cdot\mathbf{r})\mathbf{k}] - (3\omega_0^2/2R_0) \times\{[5(\mathbf{k}\cdot\mathbf{r})^2 - \mathbf{r}^2]\mathbf{k} - 2(\mathbf{k}\cdot\mathbf{r})\mathbf{r}\} + \cdots. \tag{23}$$

If only linear terms are retained in the Taylor expansion (23), the corresponding field $\mathbf{g}_L$, defined by

$$\mathbf{g}_L(\mathbf{r}) = -g_0\mathbf{k} - \omega_0^2[x\mathbf{i} + y\mathbf{j} - 2z\mathbf{k}], \tag{24}$$

will be referred to as the linearized field. Under this approximation for $\mathbf{g}$, the coordinate formulation of the g-correction equation becomes

$$\ddot{x} + \omega_0^2 x = b_x, \tag{25a}$$

$$\ddot{z} - 2\omega_0^2 z = b_z - g_0, \tag{25b}$$

with the y-equation symmetric to that in x. These differential equations, which are valid in some neighborhood of the origin, will be referred to as the linearized equations.

The linearized solution will be taken as

$$x = (v_{x,0}/\omega_0)\sin\omega_0 t + (1/\omega_0) \times \int_0^t b_x(\tau)\sin\omega_0(t-\tau)d\tau, \tag{26a}$$

$$y = (1/\omega_0)\int_0^t b_y(\tau)\sin\omega_0(t-\tau)d\tau, \tag{26b}$$

$$z = (v_{z,0}/2^{\frac{1}{2}}\omega_0)\sinh 2^{\frac{1}{2}}\omega_0 t - R_0\sinh^2\omega_0 t/2^{\frac{1}{2}} + (1/2^{\frac{1}{2}}\omega_0)\int_0^t b_z(\tau)\sinh 2^{\frac{1}{2}}\omega_0(t-\tau)d\tau. \tag{26c}$$

The case of a flat earth corresponds to $R_0\to\infty$, or equivalently $\omega_0\to 0$. Equations (26) reduce directly to the zero-order solution (11) for a flat earth as $\omega_0\to 0$ if note is taken of the limits $(\sin u)/u\to 1$, $(\sinh u)/u\to 1$ as $u\to 0$.

The linearized field $\mathbf{g}_L$ is symmetric about the z-axis. In the meridian x, z plane, the equipotentials of the linearized field are defined by

$$(z - R_0/2)^2 - x^2/2 = \text{const.}, \tag{27}$$

and the field lines are given by

$$x^2(z - R_0/2) = \text{const.} \tag{28}$$

Hence, the equipotentials form in the x, z plane a family of concentric hyperbolas and their conjugates

asymptotic to the lines $x^2=2(z-R_0/2)^2$ through the center of symmetry $(0, R_0/2)$ of the field. The linearized field diagram is shown in Fig. 5 (solid curves, field lines indicated by arrows). On the horizontal plane $z=R_0/2$, the z-component of the linearized field vanishes and above this plane the field becomes repulsive. The field diagram of Fig. 5 yields a qualitative index of the extent of the neighborhood of the origin in which the linearized equations provide a practical approximation.

In discussing the degree of approximation of the linearized solution, the error vector $\delta_L\mathbf{r}$ with components $\delta_L x$, $\delta_L y$, $\delta_L z$ will represent the true less the linearized solution. By subtracting the linearized equations, (25), from the corresponding component equations of (7), one obtains (omitting the equation for $\delta_L y$ symmetric to that in $\delta_L x$)

$$d^2\delta_L x/dt^2+\omega_0^2\delta_L x=\omega_0^2\xi x, \tag{29a}$$

$$d^2\delta_L z/dt^2-2\omega_0^2\delta_L z=\omega_0^2\zeta R_0, \tag{29b}$$

where ξ is given by (20) and the error parameter ζ is defined by

$$\zeta=1-2z/R_0-(1+z/R_0)(R_0/R)^3\sim 3(r^2-3z^2)/2R_0^2. \tag{30}$$

Physically, the right-hand sides of Eqs. (29a) and (29b) are the accelerations neglected by $\mathbf{g}_L$ on the x- and z-axes respectively. If bounds $\max(|\xi x|)$ and $\max(|\zeta R_0|)$ are known from sources other than the linearized solution itself, then one has

$$|\delta_L x|\leqslant 2\max(|\xi x|)\sin^2\omega_0 t/2, \quad (t\leqslant T_0/2); \tag{31a}$$

$$|\delta_L z|\leqslant \max(|\zeta R_0|)\sinh^2\omega_0 t/2^{\frac{1}{2}}. \tag{31b}$$

Both ξx and ζR_0 are harmonic functions, so their maximum and minimum values must be on the boundary of any closed region of space (not containing the center of the earth) considered.

It follows from (29) that the linearized x- and z-solutions are exact for motion of the box on surfaces such that $\xi x=0$ and $\zeta=0$ respectively. If one disregards the trivial locus represented by the z-axis, the reference sphere $R=R_0$ is (from Eq. (20)) the surface of zero error $\xi=0$ for the linearized x-solution (and for the linearized y-solution, by symmetry). The surface defined by $\zeta=0$ is a surface of revolution about the z-axis, and consists of two sheets intersecting the origin O_0 which are shown (dashed curves) in x, z coordinates on the linearized field diagram of Fig. 5. The upper open branch in Fig. 5 is asymptotic to the line $z=R_0/2$, and the upper and lower branches behave in the neighborhood of the origin O_0 as segments (for x positive) of the two straight lines $x^2=2z^2$. It is seen that no surface exists on which all components of the linearized solution are exact.

The leading term of the Taylor expansion of the error $\delta_L\mathbf{r}$ is given by

$$\delta_L\mathbf{r}=R_0[2v_{x,0}v_{z,0}\mathbf{i}+(v_{x,0}^2-2v_{z,0}^2)\mathbf{k}]\times(\omega_0 t)^4/4v_{e,0}^2+\cdots, \tag{32}$$

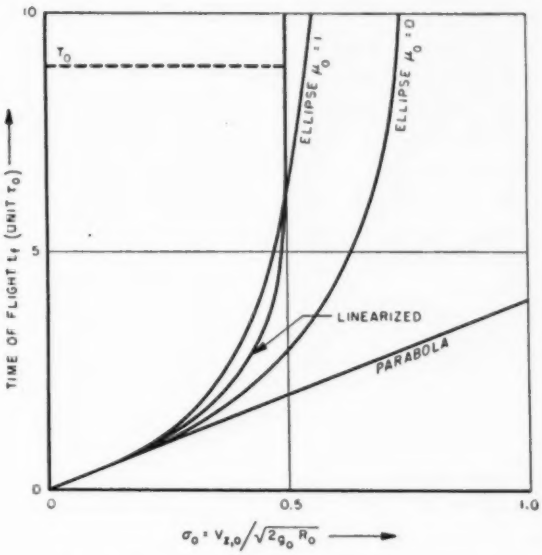


FIG. 6. Times of flight to horizontal plane $z=0$ (ballistic case, $\mathbf{b}=0$).

where $v_{e,0}$ is the escape (or parabolic) speed

$$v_{e,0}=(2g_0R_0)^{\frac{1}{2}}. \tag{33}$$

The leading terms of $\delta_L x$ and $\delta_L z$ vanish for motions initially tangent to the corresponding surfaces of zero error, respectively, and hence never vanish simultaneously. For $\mathbf{v}_0\neq 0$ the absolute error $|\delta_L\mathbf{r}|=O(t^4)$ is never of order higher than t^4, and thus the Taylor expansion of the linearized solution agrees with the true Taylor expansion of $\mathbf{r}$ through terms of order t^3, as compared to the zero-order solution (11), which agrees through terms of order t^2. Since the arc length s on the curve of motion satisfies $s=O(t)$ when $\mathbf{v}_0\neq 0$, the linearized trajectory has contact of order not less than the third with the actual trajectory at the origin, and hence the linearized solution is essentially an osculating approximation.

For the case $\mathbf{b}=0$ (ballistic trajectory), the linearized trajectory can be compared with the exact solution representing an ellipse in space and with the corresponding parabola. It is convenient to introduce the dimensionless parameters

$$\mu_0=v_{x,0}/(g_0R_0)^{\frac{1}{2}}, \quad \sigma_0=v_{z,0}/(2g_0R_0)^{\frac{1}{2}}, \tag{34}$$

for the initial velocities, where the denominators in μ_0 and σ_0 are, respectively, the satellite speed $v_{s,0}$ of (16) and the escape speed $v_{e,0}$ of (33) (one has $\sigma_0^2+\mu_0^2/2\leqslant 1$ if $v_0\leqslant v_{e,0}$). The linearized solution for $\mathbf{b}=0$ can be written

$$x=R_0\mu_0\sin\omega_0 t, \tag{35a}$$

$$z=(R_0/2)[1-(1-4\sigma_0^2)^{\frac{1}{2}}\cosh(t/\tau_0-\tanh^{-1}2\sigma_0)], \tag{35b}$$

in which the parameter τ_0 is

$$\tau_0=1/2^{\frac{1}{2}}\omega_0, \tag{36}$$

and it is assumed that $|\sigma_0|<1/2$. If $\sigma_0=1/2$, the z-solu-

tion becomes

$$z=(R_0/2)[1-e^{-t/\tau_0}], \tag{37}$$

and hence the trajectory defined by the linearized solution approaches the plane $z=R_0/2$ asymptotically as $t\to\infty$. The time constant τ_0 for z to reach the fraction $1-1/e$ of its asymptotic value in this case is approximately 9.5 min. for an initial point O_0 on the surface of the equivalent sphere S_e. For $\sigma_0>1/2$, the linearized trajectory crosses the plane $z=R_0/2$ and is unbounded in z; thus the linearized approximation implies an escape speed $v_{e,0}/2$ which is one-half the true value (33) on an inverse-square field. For $0<\sigma_0<1/2$, the time of flight t_f corresponding to intersection with the x,y plane,

$$t_f=2\tau_0 \tanh^{-1}2\sigma_0, \tag{38}$$

is compared in Fig. 6 with the corresponding time of flight $4\sigma_0\tau_0$ for a parabola, and, for a few values of μ_0, with the corresponding time of flight on an inverse-square field.

In Fig. 7, linearized trajectories from Eqs. (35) are shown (solid curves) in two families ($\sigma_0=0.3$ and $\mu_0=0.2$). From Eq. (35a) one notes that all the linearized trajectories for $\mathbf{b}=0$ intersect the z-axis at time $t=T_0/2$. For σ_0 small and thus t_f small relative to $T_0/2$ (as in the family $\sigma_0=0.3$), this intersection point is far below the x,y plane, and above this plane, the linearized trajectories show a fairly close resemblance in the large to the corresponding ellipses (dashed curves). The intersection with the z-axis is at the origin O_0 for $t_f=T_0/2$, which fixes a critical value $(\tanh\pi/2^{1/2})/2=0.488$ of σ_0 at which the upward and downward branches of the linearized trajectory above the x,y plane coincide. It is clear that for σ_0 close to $1/2$, the linearized trajectories for $\mathbf{b}=0$ are a very poor approximation in the large to the corresponding ellipses, although they are superior to the corresponding parabolas (dot-and-dashed curves) in a neighborhood of the origin.

IV. THE QUASILINEAR APPROXIMATION

The linearized approximation is based on the assumption that the gravitational acceleration $\mathbf{g}$ is a slowly varying function of position. This section discusses a linear approximation based on the additional assumptions that the radius of curvature of the trajectory is large and is a slowly varying function. The approximation will be referred to as the quasilinear approximation. For simplicity, only motion confined to the vertical x,z plane will be considered.

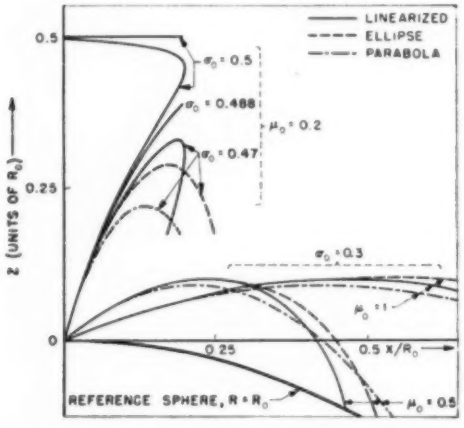


FIG. 7. Comparison of linearized trajectories with ellipses and parabolas (ballistic case, $\mathbf{b}=0$).

If the curve of motion is an analytic curve at the origin O_0, the radius vector $\mathbf{r}$ to a point on the curve can be expanded in the series[12]

$$\mathbf{r}=\boldsymbol{\alpha}_0 s+(\boldsymbol{\beta}_0/2\rho_0)s^2+\cdots, \tag{39}$$

where s is the arc length on the curve, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are the unit tangent and unit principal normal at O_0 respectively, and ρ_0 is the radius of curvature at O_0. The terms of order s^3 and higher in (39) involve first and higher derivatives with respect to s and powers above the first of the curvature $1/\rho$. If the series (39) for $\mathbf{r}$ is substituted in (23) for $\mathbf{g}(\mathbf{r})$, one obtains

$$\mathbf{g}(\mathbf{r})=-g_0\mathbf{k}-\omega_0^2[\boldsymbol{\alpha}_0-3(\boldsymbol{\alpha}_0\cdot\mathbf{k})\mathbf{k}]s - \frac{3\omega_0^2}{2}\left\{\frac{\boldsymbol{\beta}_0/3-(\boldsymbol{\beta}_0\cdot\mathbf{k})\mathbf{k}}{\rho_0} + \frac{[5(\boldsymbol{\alpha}_0\cdot\mathbf{k})^2-1]\mathbf{k}-2(\boldsymbol{\alpha}_0\cdot\mathbf{k})\boldsymbol{\alpha}_0}{R_0}\right\}s^2+\cdots, \tag{40}$$

where the terms given explicitly represent $\mathbf{g}$ under the assumption that terms of order s^3 or higher are negligible. Consider the canonical coordinates x_1 and x_2 (defined as coordinates in the direction of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ respectively) of the curve at the origin. The coordinates x_1 and x_2 are given in terms of x and z by

$$\begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}=\begin{bmatrix} \alpha_{x,0} & \alpha_{z,0} \\ \beta_{x,0} & \beta_{z,0} \end{bmatrix}\begin{Bmatrix} x \\ z \end{Bmatrix}, \tag{41}$$

where the elements of the direction-cosine matrix are components of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$; these coordinates are likewise given by

$$x_1=s, \quad x_2=s^2/2\rho_0, \tag{42}$$

from (39) under the approximation that terms of order s^3 or higher are negligible. By comparison of (41) and (42), one has the approximation

$$s=\alpha_{x,0}x+\alpha_{z,0}z, \tag{43a}$$

$$s^2=2\rho_0[\beta_{x,0}x+\beta_{z,0}z], \tag{43b}$$

which expresses both s and s^2 linearly in terms of x and z. Under the substitution (43), the terms given explicitly in (40) yield the quasilinear field $\mathbf{g}_Q$ defined (in matrix notation) by

$$\mathbf{g}_Q=-\begin{Bmatrix} 0 \\ g_0 \end{Bmatrix}-\omega_0^2\mathbf{A}_0\begin{Bmatrix} x \\ z \end{Bmatrix} \tag{44}$$

[12] W. C. Graustein, *Differential Geometry* (Macmillan Company, Inc., New York, 1947), p. 39.

where the coefficient matrix $\mathbf{A}_0$ is

$$\mathbf{A}_0=\begin{bmatrix}1-6\alpha_{x,0}\alpha_{z,0}\beta_{x,0}\rho_0/R_0 & 6\alpha_{x,0}\alpha_{z,0}\beta_{z,0}\rho_0/R_0\\ 3\beta_{x,0}(3\alpha_{z,0}{}^2-1)\rho_0/R_0 & -2+3\beta_{z,0}(3\alpha_{z,0}{}^2-1)\rho_0/R_0\end{bmatrix}. \tag{45}$$

The parameters entering the coefficient matrix $\mathbf{A}_0$ can be evaluated directly from the initial conditions of the motion. From standard formulas,[12] one has

$$\boldsymbol{\alpha}_0=\frac{\mathbf{v}_0}{v_0},\quad \boldsymbol{\beta}_0=\frac{\mathbf{v}_0\times(\mathbf{b}_0-g_0\mathbf{k})}{|\mathbf{v}_0\times(\mathbf{b}_0-g_0\mathbf{k})|}\times\frac{\mathbf{v}_0}{v_0}, \tag{46}$$

and

$$\rho_0=v_0{}^3/|\mathbf{v}_0\times(\mathbf{b}_0-g_0\mathbf{k})|, \tag{47}$$

where $\mathbf{b}_0$ is the initial value of $\mathbf{b}$ and it is assumed that $\mathbf{v}_0$ is non-vanishing. The corresponding quasilinear equations of motion are given by

$$\begin{Bmatrix}\ddot{x}\\ \ddot{z}\end{Bmatrix}+\omega_0{}^2\mathbf{A}_0\begin{Bmatrix}x\\ z\end{Bmatrix}=\begin{Bmatrix}b_x\\ b_z-g_0\end{Bmatrix}. \tag{48}$$

Solutions of the system (48) can be obtained directly by the use of the Laplace transform.

The quasilinear approximation represented by Eqs. (48) yields a transition from the linearized approximation (25) to the spherical approximation (14). The linearized equations assume that terms of the type $(x/R_0)^2$ in the Taylor expansion of $\mathbf{g}$ are negligible compared to terms of type x/R_0, which is equivalent to assuming $(s/R_0)^2\ll s/R_0$. From Eqs. (43) this condition requires that $\rho_0\ll R_0$, and one notes that Eqs. (48) reduce to the linearized equations (25) when $\rho_0\to 0$. The spherical solution (15) for motion in the x,z plane presupposes $\boldsymbol{\alpha}_0=\mathbf{i}$ and $\boldsymbol{\beta}_0=-\mathbf{k}$ at the origin; if note is taken of the constraining condition (12) on $\mathbf{b}_0$, Eq. (47) yields $\rho_0=R_0$. With these values of the parameters, Eqs. (48) reduce directly to the spherical equation, (14), of motion.

Since the terms in the Taylor expansion of $\mathbf{g}$ neglected by $\mathbf{g}_Q$ are of order s^3, and $s=O(t)$ when $\mathbf{v}_0\neq 0$, it follows that the absolute error $|\delta_Q\mathbf{r}|$ of the quasilinear solution is such that $|\delta_Q\mathbf{r}|=O(t^5)$ in general. Thus, the Taylor expansion of the quasilinear solution agrees with the true Taylor expansion of $\mathbf{r}$ through terms of order t^4 in general, as compared to the linearized solution (26), which agrees through terms of order t^3 at most. The quasilinear trajectory has contact of order not less than the fourth with the actual trajectory at the origin, and accordingly is an osculating approximation in general.

One can define error parameters for the quasilinear equations analogous to ξ and ζ in the linearized case, which fix surfaces of zero error for the component equations of (48). By this means one can show that the surfaces of zero error for both components coincide, and the quasilinear system (48) is exact if and only if the motion is constrained to the reference sphere S_0 (condition (12) holds), in which case the quasilinear system reduces to the spherical equation, (14). When $\rho_0\sim R_0$, the surfaces of zero error for both components are nearly in coincidence, and the quasilinear solution provides an approximation which is markedly superior to the linearized. When ρ_0 is small relative to R_0, the quasilinear trajectory differs only slightly from the linearized, and may yield only a small margin of gain. These results are illustrated by Fig. 8, which compares the quasilinear trajectories for the ballistic case ($\mathbf{b}=0$) with the corresponding ellipses and linearized trajectories when $\sigma_0=0$.

In passing, it can be noted that generalization of this quasilinear approximation to three dimensions is easily made (by assuming that the torsion of the curve of motion is likewise a slowly varying function).

V. LINEAR CONTINUATION

When significant variations of altitude above the earth occur on a trajectory of extended range, it is clear that none of the linear approximations given for $\mathbf{g}(\mathbf{r})$ are satisfactory in the large. In this case, the g-correction equation, (7), can always be solved by a step-by-step procedure in which the non-linear gravitational term over any interval of time is replaced by a constant value corresponding to the interval, as determined from the solution of the equation for times up to the interval in question. The solution in any interval is given by a suitable modification of the zero-order solution (11), and the solution is continued from interval to interval by identifying the initial conditions of one interval with the terminal conditions of the preceding interval. Such a method will be referred to as zero-order continuation, and, while always applicable, has the disadvantage of requiring a large number of steps for an extended range.

To reduce the number of steps required, the gravitational acceleration over any interval of time can be

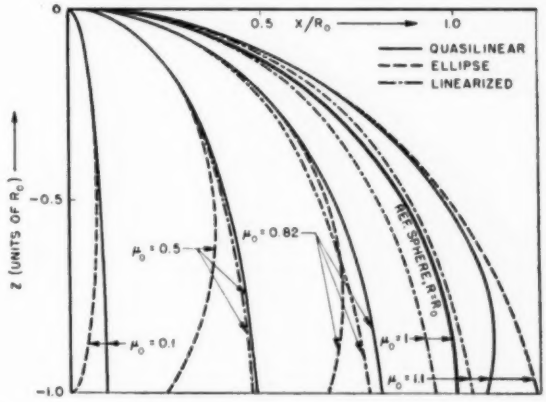


FIG. 8. Comparison of quasilinear trajectories with ellipses and linearized trajectories (ballistic case, $\mathbf{b}=0$).

TABLE I. Results of continuation solutions.

Solution	Step solution	Number of steps	Error $\|\delta\mathbf{r}\|$ (mi.)
I	Linearized and quasilinear*	5	20
II	Linearized	39	46
III	Zero-order	39	254

* As specified in Fig. 3.

replaced by a linear approximation whose constants are determined by the solution up to the interval. The most advantageous approximation is the quasilinear. The initial conditions for any interval are again taken as the computed terminal conditions of the preceding interval. Such a method of solution will be referred to as linear continuation.

For simplicity, let the motion be plane with an initial point O_0 on the surface of the equivalent sphere S_e, and let the components of $\mathbf{b}$ be measured parallel to the axes x, z of the frame O_0. Let the total range, O to t_n of t, be divided into n intervals, t_{k-1} to t_k $(k=1, \cdots, n)$ with $t_0=0$. The computed terminal point O_{k-1} in the $(k-1)$th interval determines a frame $x_{(k)}$, $z_{(k)}$ with parameters R_{k-1}, Θ_{k-1} for the kth interval, as in Fig. 1 $(\Theta_0=0)$. In the frame O_{k-1} (which is introduced for computational convenience), the theory of the preceding sections for the frame O_0 can be applied to define the pertinent parameters g_{k-1}, ω_{k-1}, $\mathbf{A}_{k-1}$ by interchange of subscripts. If $\mathbf{B}_{k-1}$ designates the direction-cosine matrix

$$\mathbf{B}_{k-1}=\begin{bmatrix} \cos\Theta_{k-1} & \sin\Theta_{k-1} \\ -\sin\Theta_{k-1} & \cos\Theta_{k-1} \end{bmatrix}, \tag{49}$$

the g-correction equation in the kth interval when referred to the O_{k-1} frame is

$$\begin{Bmatrix} \ddot{x}_{(k)} \\ \ddot{z}_{(k)} \end{Bmatrix} + \omega_{k-1}{}^2\mathbf{A}_{k-1}\begin{Bmatrix} x_{(k)} \\ z_{(k)} \end{Bmatrix} = \mathbf{B}_{k-1}{}'\begin{Bmatrix} b_x \\ b_z \end{Bmatrix} - \begin{Bmatrix} 0 \\ g_{k-1} \end{Bmatrix}, \tag{50}$$

where $\mathbf{B}_{k-1}{}'$ is the transpose of $\mathbf{B}_{k-1}$. The formulation (50) corresponds to the use of a quasilinear approximation, which reduces to a linearized or spherical approximation in the special cases

$$\mathbf{A}_{k-1}=\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}, \quad \mathbf{A}_{k-1}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{51}$$

respectively. The g-correction equation in the kth interval is then

$$\begin{Bmatrix} \ddot{x} \\ \ddot{z} \end{Bmatrix} + \omega_{k-1}{}^2\mathbf{B}_{k-1}\mathbf{A}_{k-1}\mathbf{B}_{k-1}{}'\begin{Bmatrix} x \\ z \end{Bmatrix} = \begin{Bmatrix} b_x \\ b_z \end{Bmatrix} - \mathbf{B}_{k-1}\begin{Bmatrix} 0 \\ g_{k-1} \end{Bmatrix} \tag{52}$$

when referred to the O_0 frame.

In practice, it is found that linear continuation yields a considerable reduction in number of intervals as compared with zero-order continuation for the same terminal error. This fact appears by comparison of solutions I and II with solution III in Table I, which shows the results of three different continuation solutions carried out for the A10 trajectory. Solution I was based on the use of quasilinear approximations in the intervals (as specified in Fig. 3) where $\rho\sim R_0$; comparison of the results with those of solution II brings out the superiority of the quasilinear over the linearized approximation under this condition. In carrying out these computations, the $\mathbf{b}$-function (Fig. 4) was tabulated to four decimal places (after round-off) in units of g_e from the assumed trajectory (consisting of segments of analytic curves). Checks indicated that the rounding-off process in $\mathbf{b}$ introduced an error $|\delta\mathbf{r}|$ which amounted to <3 mi. for the solutions of Table I, which shows that the entries in the table for $|\delta\mathbf{r}|$ are significant for comparison of the approximations to $\mathbf{g}$.

It should be emphasized that the computations yielding Table I are of academic nature, due to neglect of the latitude-dependent component of $\mathbf{g}$ and of the centrifugal and Coriolis terms in the kinetic reaction of the g-correction equation. The sum of the first two of these acceleration terms varies over a range of $0.005g_e$ from the equator to the pole on the earth's surface. The systematic errors due to neglect of the variation of these terms on the A10 trajectory have been calculated approximately by a perturbation method for the case when the plane of motion coincides with the earth's equatorial plane (the error is then in this plane likewise). In this case the effect of variation of centrifugal reaction ($|\delta\mathbf{r}|\sim 30$ mi.) and of Coriolis reaction ($|\delta\mathbf{r}|\sim 700$ mi.) is extreme. These errors are larger by orders of magnitude than the upper limit ($|\delta\mathbf{r}|<3$ mi.) to the round-off error found above for the data of Table I. Hence, the method of computation leading to Table I, in the particular form used, has only a limited practical validity for the A10 trajectory (except as an approximation in the roughly constant plane of motion for a trajectory in the neighborhood of a pole, where the Coriolis deflection is transverse).

It is clear from these considerations that the step solution in linear continuation must be modified on trajectories of extended range, in general, to include terms corresponding to the variation of the latitude-dependent component of $\mathbf{g}$, the variation of the centrifugal reaction, and the Coriolis reaction. The fact that these terms are considerably smaller than the main radial term of $\mathbf{g}$ materially simplifies the problem, which will not be attempted here because of the detail involved.

The advantage of linear continuation over numerical methods such as the Runge-Kutta process consists in the fact that the length of step or the integration interval is fixed by the variation of the slowly varying $\mathbf{g}$-function. From such error bounds as (22) or (31), one can estimate the number of steps required for a given terminal accuracy. In methods of the Runge-Kutta type, however, the integration interval is fixed by the

variation of the rapidly varying **b**-term, and the possible interval is considerably smaller. Linear continuation yields a reasonable compromise between the conflicting demands of a small number of steps and simplicity of the step solution.

In conclusion the authors wish to acknowledge many helpful discussions of this problem with Dr. W. C. Randels of Northrop Aircraft Corporation, Dr. R. Isaacs of The Rand Corporation, and Prof. T. Dantzig of the Rand consulting staff. Thanks are due also to Mrs. M. Irving and Mrs. J. Griffith for the computational work.

Diffraction Pattern in a Circular Aperture Measured in the Microwave Region

C. L. Andrews
General Electric Research Laboratory, The Knolls, Schenectady, New York
(Received February 2, 1950)

Measurements were made of the diffraction patterns of circular apertures from one to eight wave-lengths in diameter in the planes of the apertures and in the neighborhood of the apertures when a plane polarized electromagnetic wave was incident upon them. From Thomas Young's theory that the diffraction pattern is an interference pattern between the incident plane wave and wavelets from the edge of the aperture, the positions of maximum intensity have been predicted in the neighborhood of the aperture and the values of the intensities over the apertures checked with experiment.

INTRODUCTION

DIFFRACTION of light waves has been explained by Thomas Young as interference between the geometrically propagated incident wave and Huygen's wavelets caused by disturbance at the edge of the diffracting screen. Young's explanation was dropped promptly for Fresnel's theory that the diffraction pattern of a plane screen is the resultant of interfering wavelets from every point in any open portion of the infinitely extended plane of the screen. Lacking experimental facts, Fresnel made the simple assumption that the wave reaching the aperture of the screen was unperturbed. On the basis of Green's theorem and Fresnel's assumptions, Kirchoff developed a more exact theory of diffraction. A hundred years after Young and Fresnel, Rubinowicz[1] showed that Kirchoff's surface integral for the determination of the disturbance at a point on the opposite side of the aperture from the source may be transformed into two terms, one for the geometrically propagated wave through the aperture, and the other a line integral around the edge of the aperture. Thus Rubinowicz gave us two mathematically equivalent ways of thinking of diffraction by an aperture, one to think of intensity at a point in the diffraction pattern as the sum of all the effects from every point in the aperture and the other to think of diffraction as an edge effect.

Recently microwaves by their convenient length have provided a means of measuring the diffraction pattern in the plane of an aperture and on the sides toward and away from the source. It has been emphasized that the sharpest part of the diffraction pattern of an aperture is in the plane of the aperture itself.[2] Although this fact was recognized through measurements of microwave intensities in apertures, it could have been observed with light.

If the elementary apparatus for study of diffraction of light by a single slit is employed, the aperture may be opened to uncover a half dozen Fresnel strips relative to the point of observation. If the eye piece is then moved slowly toward the slit and at the same time the slit is diminished in width, so that the same number of dark bands is always observed, then, when the eye piece is finally focused on the plane of the aperture, sharp black lines are seen, each separated by a wavelength.

These simple observations with light are mentioned that the reader may confirm a fact of diffraction that has not been noted in the textbooks. The experimental work reported in this paper was performed with microwaves which provide the most precise means of measuring the diffraction patterns near apertures of the order of a few wave-lengths in diameter.

The assumption by Fresnel of constant intensity across the aperture, or more recent assumptions of constant intensity over most of a large aperture with tapering off only at the edge, have justification only in that they may be used in calculating correct intensities in the diffraction pattern beyond the aperture. The assumptions do not agree with experiment.

Thomas Young's theory of diffraction correctly predicts the positions of maximum and minimum intensity on the sides of the aperture toward and away from the source and in the plane of the aperture itself. Young assumed that the diffraction pattern of an aperture is the interference pattern between a geometrically propagated wave through the aperture and reradiated wavelets from the edge of the aperture. Thus the diffraction pattern of a slit is simply the interference pattern of three waves, the geometrically propagated wave through

[1] Rubinowicz, Ann. d. Phys. **53**, 257 (1917).
[2] C. L. Andrews, Phys. Rev. **71**, 777 (1947).

the slit and two cylindrical waves from the edges. The interference pattern of a plane wave and a cylindrical wave is a family of parabolas with their common principal focus at the edge, and their focal lengths an odd number of quarter wave-lengths. The interference pattern of the three waves in the plane perpendicular to the edges is not a set of curves but an array of points where the two families of parabolas intersect as shown in Fig. 1, the points where all three waves are in phase. Where the parabolas intersect at small angles the contours of intensity are elongated loops but near the aperture where the parabolas intersect nearly perpendicularly the contours are nearly circular and the gradients of intensity largest. Figure 1 gives the points of maximum intensity in the diffraction pattern of a slit three wave-lengths wide. Whether points A and B be positions of two long rods or the edges of a slit or the edges of a circular aperture upon which a plane polarized wave is incident, this simple theory of Thomas Young predicts closely the positions of maximum intensity in the magnetic plane between two parallel lines from the source through A and B. As the slit edges are separated the two families of parabolas move with them while the peaks of intensity move forward through the aperture from the standing wave region to the Fresnel region.

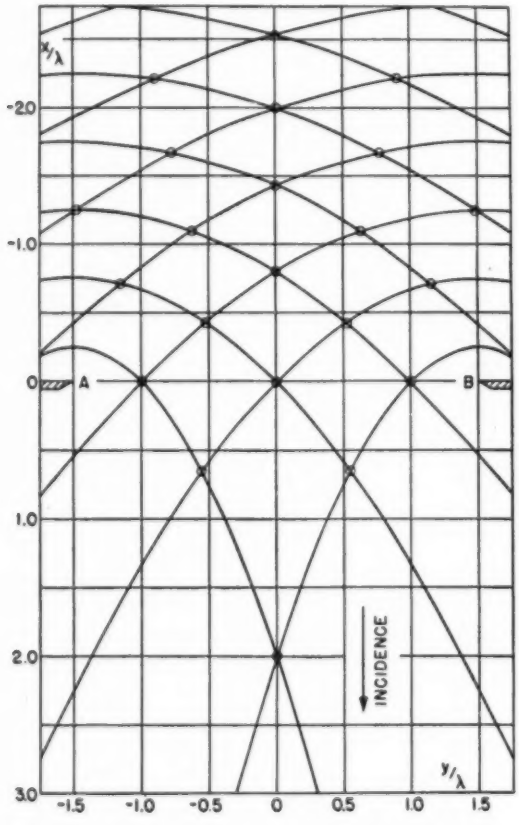


FIG. 1. Diagram for determination of positions of maximum intensity in the diffraction pattern in the magnetic plane of a circular aperture when a plane polarized wave is incident normally.

APPARATUS

The experimental arrangement was similar to that described previously.[1] The wave-length was reduced from 12.8 cm to 8.0 cm. A special triode oscillator was built for the purpose.[3] The parabolic reflector was 4 ft. (122 cm) in diameter. The equipment was set up out-of-doors on a roof to avoid reflections from walls and the distance from source to diffracting screen increased to 75 ft. (23 meters). The centers of the parabolic reflector and diffracting screen were 6 ft. above the roof. The block of the optical bench was driven by motor and screw at the rate of two wave-lengths per minute and the intensities recorded on a recording galvanometer (General Electric Photoelectric Recorder) with a sensitivity of 10 microamperes full scale and a period of 0.83 second.

There were two major sources of error (1) the lack of constancy of intensity in the cross section of the incident unperturbed beam, (2) the finite size of the antenna of the probe and the leads from the probe to the recorder. The automatic recorder made the search for sources of error simpler. The greatest variations of intensity of the unperturbed beam over a width of four wave-lengths were four percent and over a width of eight wave-lengths 10 percent. The crystal itself, Sylvania IN23B, with the brass tip of its capsule removed was used as dipole so that no coaxial lines would have to be employed to

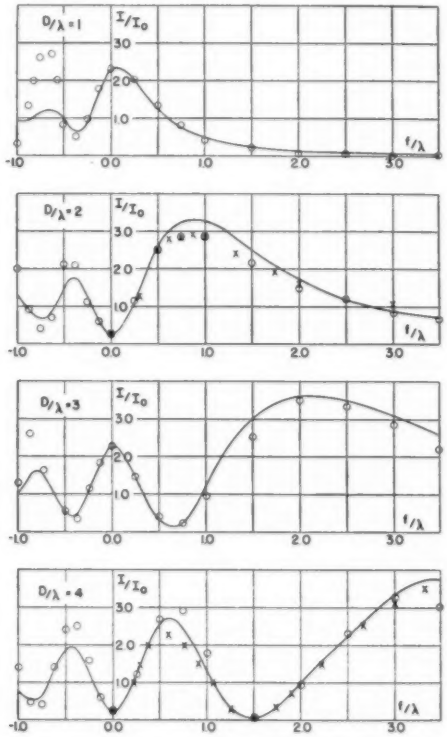


FIG. 2. Diffraction patterns along axes of circular apertures 1, 2, 3, and 4 wave-lengths in diameter as computed from Kirchoff's theory. Circles are experimental data. Crosses are data of Severin.

[3] C. L. Andrews, General Electric Review **50**, 40 (1947).

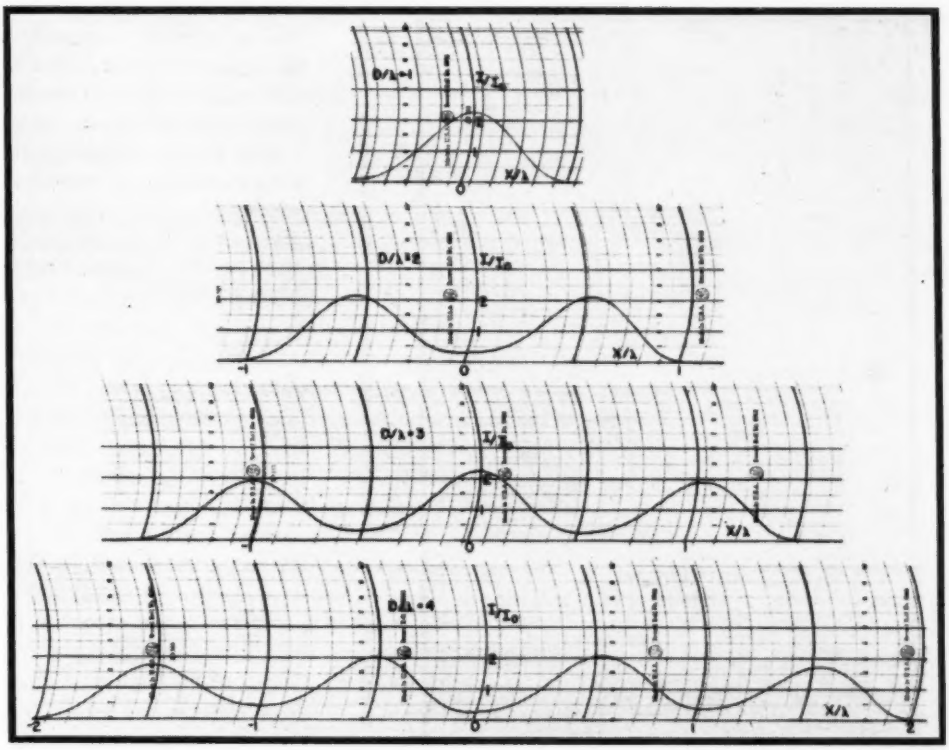


FIG. 3. Recorded graphs of intensity of radiation along magnetic diameters of circular apertures 1, 2, 3, and 4 wave-lengths in diameter.

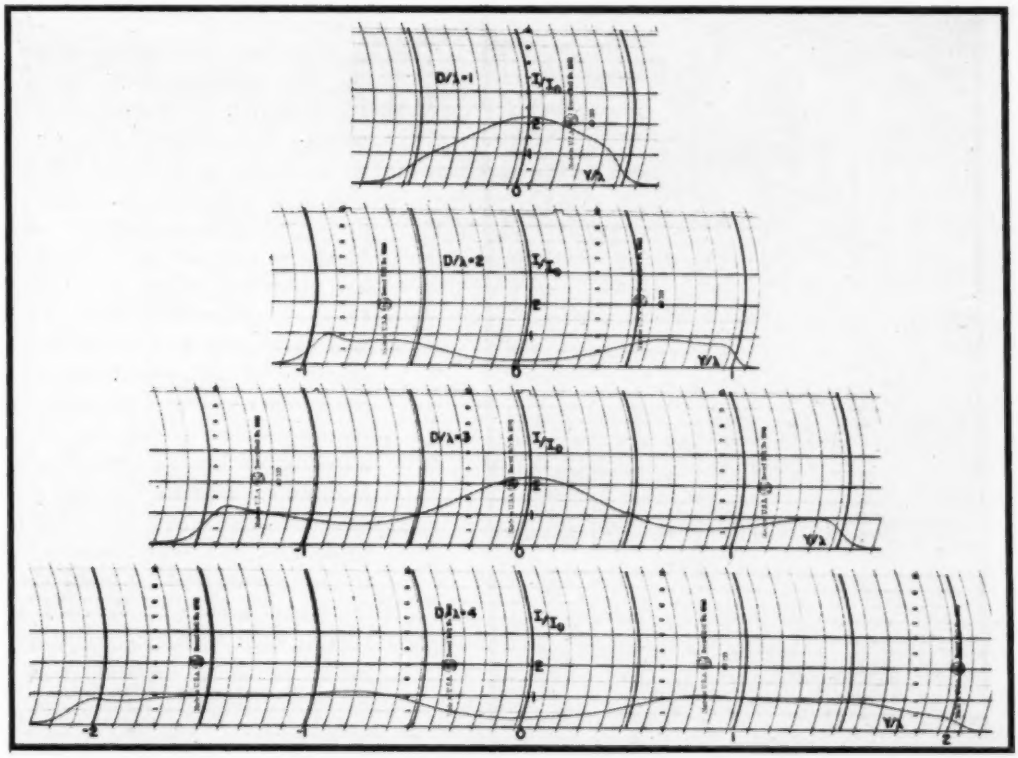


FIG. 4. Recorded graphs of intensity of radiation along electric diameters of circular apertures 1, 2, 3, and 4 wave-lengths in diameter.

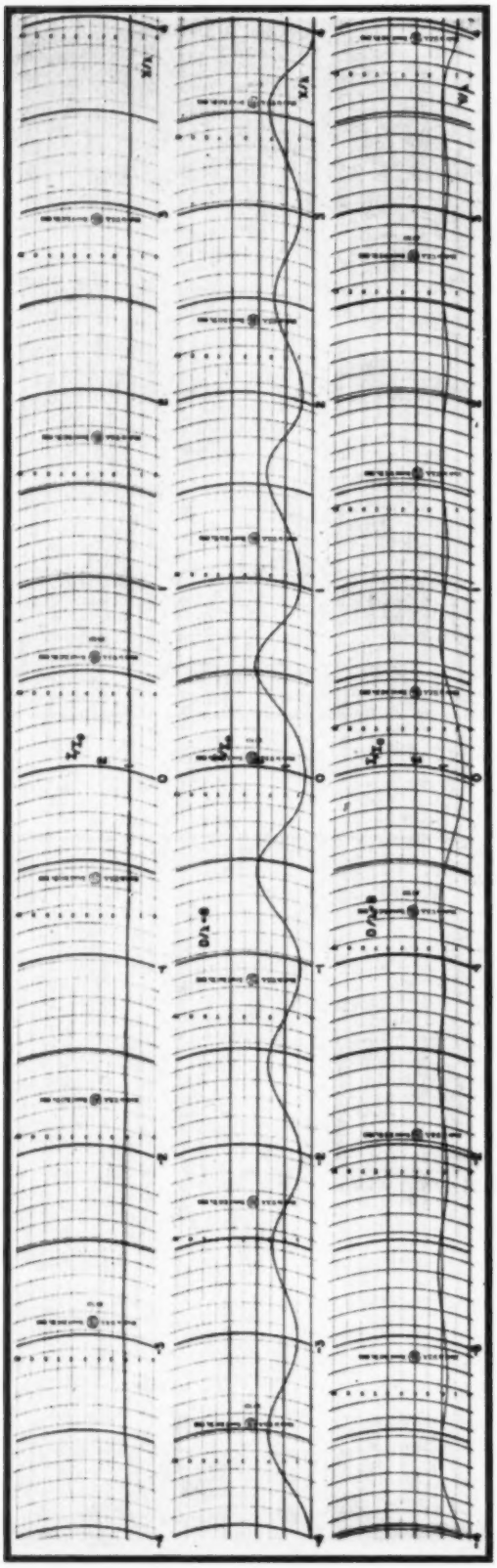


FIG. 5. Intensity of radiation of the unperturbed beam and along magnetic and electric diameters of circular apertures eight wave-lengths in diameter.

disturb the field. A fine twisted lead of No. 23 wire was attached with soft solder to the ends of the capsule of the sealed silicon crystal and extended in the direction of propagation for three meters and guyed by strings and rubber bands. Thence it extended out of the beam perpendicularly to the electric field.

The errors caused by the finite size of the antenna were estimated in two ways:

(1) By comparing records of the same diffraction patterns as the length of the probe was reduced.

(2) By studying the diffraction patterns of rods varying from 0.5 to 0.1 wave-length.

When the antenna was reduced from 0.25 to 0.20 wave-length, the patterns were not altered noticeably except when the probe was closer than a quarter wave-length to the edge of an aperture. Likewise, it was noted that rods of 0.2 wave-length long had negligible effects on the field at distances greater than one-quarter wave-length.

It was estimated that the greatest source of error in all patterns shown in the following graphs was the variation of intensity in the cross section of the unperturbed beam except at points within an eighth wave-length of an edge on the H diameter or a quarter wave-length of an edge on the E diameter where the largest error was caused by the finite size of the probe.

EXPERIMENTAL RESULTS

In Fig. 2 the circles indicate experimental data of intensity of radiation I relative to the intensity I_0 of the unperturbed beams taken along the axes of circular apertures in metal screens 1, 2, 3, and 4 wave-lengths in diameter. f/λ is the distance in wave-lengths along the axis measured from the plane of the aperture in the direction of propagation. The crosses are experimental data taken from the results of Severin[5] for wave-lengths of 6 and 10 cm. Severin coated the front and back of his metal screen with a "paint" that reduced reflections to 10 percent of the metallic reflection. Near the aperture data are the same for the reflecting and absorbing screen. The curves of Fig. 2 are plots of the calculations from Kirchoff's theory.[6] Note that the best agreement with experiment is at the center of the aperture. If the assumptions of Kirchoff's theory were not conflicting and if the solution were correct then the solution should yield the original assumption of an unperturbed beam in the plane of the aperture. Therefore, we shall ignore the origins of Kirchoff's expression and consider it as a semi-empirical formula. The disagreement with experimental data on the side toward the source may be expected because of reflection from the mirror surface on the back side of the screen.

Note that when the aperture is an odd number of wave-lengths in diameter I/I_0 at the center is 2.25. When the diameter is an even number of wave-lengths, I/I_0 is 0.25.

[5] H. Severin, Zeits. f. Naturforshung, I 9, 487–495 (1946).
[6] See reference 1, Eq. (4) and Fig. 1.

Figure 3 is of recorded graphs of the intensity along the magnetic diameters of circular apertures 1, 2, 3, and 4 wave-lengths in diameter when a plane polarized beam is incident normally on the screen.

Figure 4 is a similar record of intensities along the electric diameters. Note that the measurements along the electric diameters are not dependable closer than a quarter wave-length from the edge. It may be noted that whatever the length of the dipole probe the recorded intensity begins to drop as soon as one end of the dipole probe goes behind the screen and drops to zero when the probe is completely hidden from the source.

Figure 5 is a recording of the intensity of the unperturbed beam and the intensities along the magnetic and electric diameters of an aperture eight wave-lengths in diameter.

A search for a meaningful empirical expression to describe the patterns resulted in

$$(u/u_0)_v = 1 - \frac{1}{2\pi} \oint \epsilon^{j\beta} \cos\theta \cos\gamma (ds/r), \qquad (1)$$

$$(u/u_0)_h = -\frac{1}{2\pi} \oint \epsilon^{j\beta} \cos\theta \sin\gamma (ds/r), \qquad (2)$$

u is the complex amplitude of the resulting wave at point P of Fig. 6 in terms of u_0, the amplitude of the incident wave. The subscripts v and h indicate vertical and horizontal components. The term unity is for the vertically polarized incident beam. The negative sign indicates the reversal of phase of the reradiated wavelet from the edge. ds is an element of edge and r its distance from point P. β is the phase angle, $2\pi r/\lambda$. Multiplication by $\cos\theta$ resolves the component of the amplitude of the incident wave parallel to the element of edge ds. $\cos\gamma$ and $\sin\gamma$ resolve the amplitude of the wavelet from ds arriving at P into vertical and horizontal components. Since the expression contains ideas clearly expressed by Thomas Young, we shall call it Young's Circuital Form for determining the diffraction pattern in the plane of an aperture in a flat screen when a plane polarized transverse wave is incident normally on the plane of the aperture.

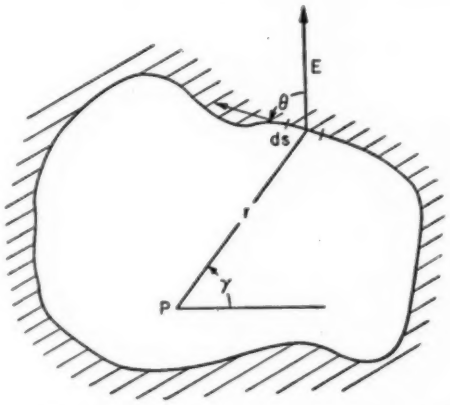


FIG. 6. Aperture in screen upon which a vertically polarized beam is incident normally.

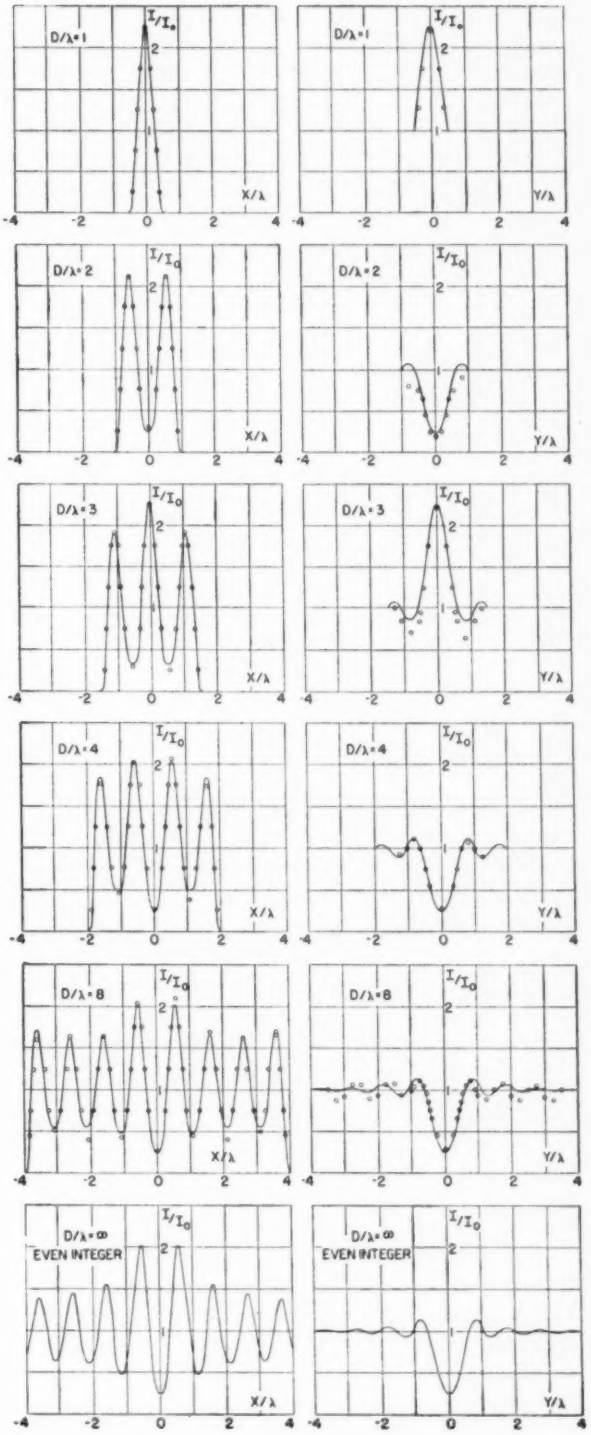


FIG. 7. Computed intensities of radiation along magnetic and electric diameters of circular apertures 1, 2, 3, 4, 8 and an infinite even integer number of wave-lengths in diameter. The circles indicate experimental data from Figs. 3, 4, and 5.

If the aperture is a circle of radius R and x and y are

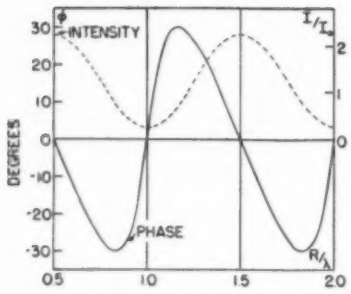


FIG. 8. Phase and intensity of radiation at center of circular aperture *versus* diameter.

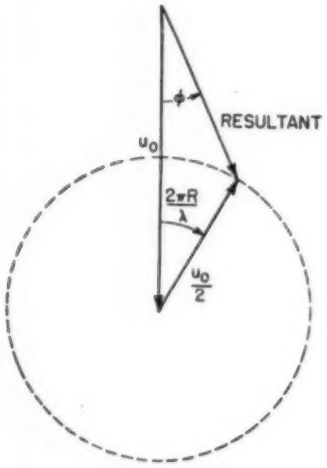


FIG. 9. Vector diagram for determination of phase and amplitude at center of circular aperture one wave-length in diameter or greater.

the cartesian coordinates taken along the horizontal and vertical diameters from the center to express the position P, then

$$(u/u_0)_v = 1 - \frac{1}{2\pi}\oint \epsilon^{i\beta}(R/r)^2(\cos^2\theta - (x/R)\cos\theta)d\theta, \quad (3)$$

$$(u/u_0)_h = -\frac{1}{2\pi}\oint \epsilon^{i\beta}(R/r)^2 \times(\sin\theta\cos\theta - (y/R)\cos\theta)d\theta. \quad (4)$$

All of the independent variables may be expressed in terms of any one of them. We were not able to integrate the expressions but evaluated the integrals for points along the magnetic and electric axes by dividing the circle into ten degree steps and adding vectorially. Some cases were checked by dividing the aperture into two degree steps.

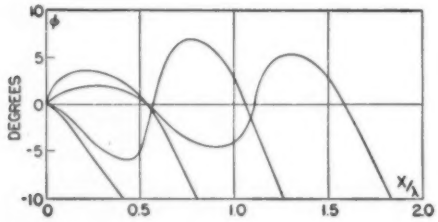
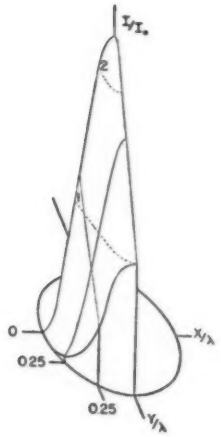


FIG. 10. Phases along the magnetic diameter of apertures 1, 2, 3, and 4 wave-lengths in diameter.



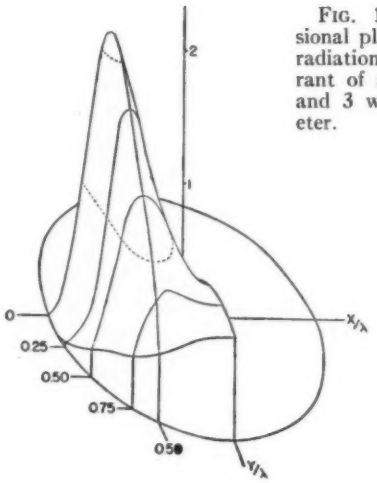
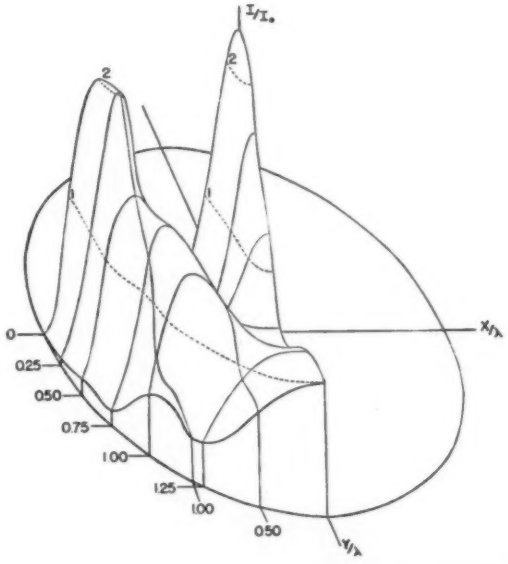


FIG. 11. Three dimensional plots of intensity of radiation over one quadrant of apertures of 1, 2, and 3 wave-lengths diameter.



The two columns of curves in Fig. 7 are of the computed intensities of radiation along magnetic and electric diameters of circular apertures 1, 2, 3, 4, 8 and an infinite even integer number of wave-lengths in diameter. The circles indicate experimental data from the curves of Figs. 3, 4, and 5.

Two features other than the general agreement of the data and calculations are to be noted. For both the experimental and calculated curves (1) the peaks are narrower than the valleys (2) the maxima and minima do not lie at exactly integer half wave-lengths from the center but slightly beyond those positions.

At the center of the circular aperture the phase ϕ alternates with increasing radius between 30° and −30° by a saw-toothed relation indicated by Fig. 8. A plot of I/I_0 *versus* R/λ is given with coordinates on the right to indicate that, when the intensity differs most from that of the incident beam, the phase is the same as that of the incident beam.

At the center

$$I/I_0 = 5/4 - \cos 2\pi R/\lambda, \tag{5}$$

$$\tan\phi = \frac{\frac{1}{2}\sin 2\pi R/\lambda}{1-\frac{1}{2}\cos 2\pi R/\lambda}. \tag{6}$$

Equations (5) and (6) are seen to follow from the vector diagram of Fig. 9.

Equations (5) and (6) are a special case of Kirchoff's expression[6] for the amplitude and phase on the axis provided $R/\lambda > \frac{1}{2}$. For smaller radii the point of observation may be chosen on the axis near the center on the side of the screen away from the source. Small apertures have not been treated experimentally. They will require a smaller probe than the cases reported here.

Figure 10 is a plot of phases along the H diameter for diameters of 1, 2, 3, and 4 wave-lengths. The calculations of these angles may be in error by as much as 0.5 degree. The phase angles are small and where the intensity is a maximum the phase is the same as that of the incident beam. The largest phase angles occur where the intensity is approaching zero at the edge. Thus with good approximation one may assume constant phase over the aperture when applying Huygens' principle to calculate the diffraction pattern beyond the aperture. Smythe's[7] method of the double current sheet will be readily applicable. However, the approximation of constant phase is justified only when the diameter of the aperture is an integer number of wave-lengths.

If, for apertures an integer number of wave-lengths wide, one considers the positions of maximum intensity, Fig. 8, as line sources, then the Fraunhaufer diffraction pattern is easily calculated vectorially and is the same as that calculated by zone or strip methods from the assumption of constant phase and intensity over the aperture.

Figure 11 is of three-dimensional plots of intensity over one quadrant of apertures of one, two and three wave-lengths in diameter from experimental data.

CONCLUSION

In view of these experimental observations the classical treatment of diffraction is in need of review. Although there have been many volumes of theory written on diffraction, there have been comparatively few experimental results reported in sixty years. Microwaves now provide a precise means of measurement of diffraction patterns near apertures.

Young's theory of diffraction yields the diffraction pattern in the plane of a circular aperture.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the support and helpful discussions throughout this study of W. C. White and E. D. McArthur of the General Electric Research Laboratory.

[7] W. R. Smythe, Phys. Rev. **72**, 1066 (1947).

The Retrograde Motion of the Arc Cathode Spot

C. J. GALLAGHER
General Electric Research Laboratory, Schenectady, New York
(Received February 6, 1950)

The motion of the cathode spot in a direction opposite to that predicted by Ampère's law depends on the arc current, gas pressure, kind of gas, and magnetic field strength. Studies of the effects of these variables have been made, including measurements of velocity and the critical pressure at which reversal of motion occurs. The phenomena observed have not been clearly explained by any of the pictures presented to date. The existence of the retrograde motion indicates very strongly that the positive-ion space charge outside the cathode is all important in determining the mechanism of current transfer.

IN the low pressure arc the cathode spot, when subjected to a transverse magnetic field, may move in a direction opposite to that which Ampère's law predicts.[1,2] The direction and speed of this retrograde motion are determined by: (1) gas pressure, (2) magnetic field strength, (3) arc current, and (4) kind of gas. It is not certain at this writing whether or not the cathode material has any appreciable effect. This paper will describe some experiments in which the above variables were investigated as to their effect on the motion.

The first series of experiments is concerned with a low pressure mercury arc, using a mercury pool cathode. The geometry of the tubes used was such that the cathode was ring-shaped. The tube was placed in the gap of an electromagnet with an annular structure as shown in Fig. 1. The magnetic field was radial, so that the current entering the cathode spot was always perpendicular to the field. Both tube and magnet were immersed in oil, which was cooled by circulating through a heat exchanger to provide a steady stream of oil cooler than the bath itself. This stream was directed against the wall of the tube and thus determined the condensing temperature of the mercury inside. The condensing temperature could be considered the same as that of the stream, except for the small thermal drop in the glass wall of the tube. The tube was operated at a line voltage of 250 v d.c. with a variable series resistance to regulate the current.

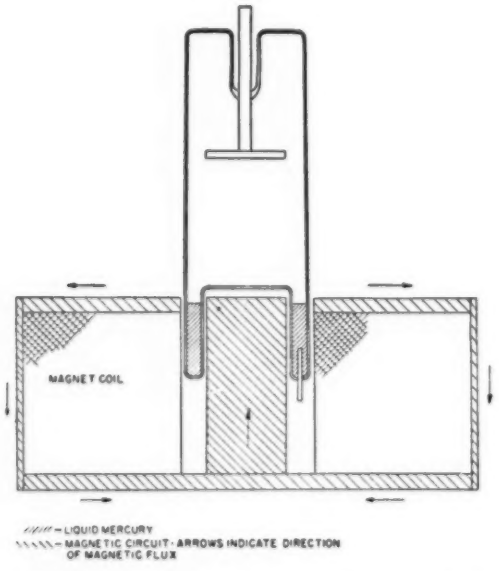


FIG. 1. Mercury pool tube in radial magnetic field.

The velocity of the spot was obtained from the frequency of rotation. The light from the region of the spot was transmitted through the oil by a glass rod and then allowed to fall on a photo-multiplier tube, the output of which was applied to the vertical plates of an oscilloscope. A sine wave of known frequency was applied to the horizontal plates and varied until a one to one Lissajou pattern was obtained. The method was not effective below about 20 c.p.s. because of random fluctuations in the position of the path of the spot over the mercury surface. At high speeds, the spot stayed against the wall of the tube so that very regular patterns could be obtained.

The velocity of the spot as a function of magnetic field, arc current, and pressure is shown in Figs. 2, 3, and 4, respectively. As shown in Fig. 2, the velocity varies almost linearly with magnetic field over the middle range, with an exponential-like variation at low fields. The slopes of the linear portions of the curves appear to be independent of current. At high fields, above 3000 gauss, the results are somewhat uncertain. It was found that the points could not be reproduced with any great accuracy. The voltage across the arc at high magnetic fields is high and may cause instability, although the arc voltage did not show any unusual fluctuations in this region such as might be expected if the arc were unstable. In Fig. 3, the same data are replotted, but with arc current as the independent variable. There is a definite leveling-off point for each value of field, with the constant velocity region setting in at smaller values of current as the field is increased.

The velocity of the spot depends on the pressure, decreasing as the pressure increases. In fact, if the pressure is sufficiently increased, the spot can be made to slow down to zero velocity and then reverse its motion. The pressure at which reversal occurs depends on the magnetic field and the current, as shown in Fig. 4. The reversing pressure as a function of magnetic field is shown for different arc currents. In general, less field

[1] N. Minorsky, J. de phys. et rad. **9**, 127 (1928).
[2] C. G. Smith, Phys. Rev. **62**, 48 (1942).

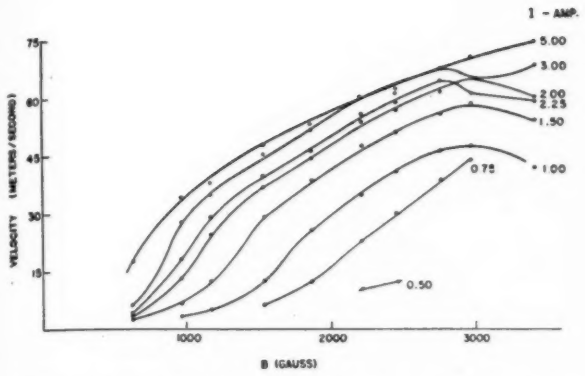


FIG. 2. Velocity of cathode spot on Hg pool *vs.* magnetic field for various arc currents. Pressure = 0.7μ.

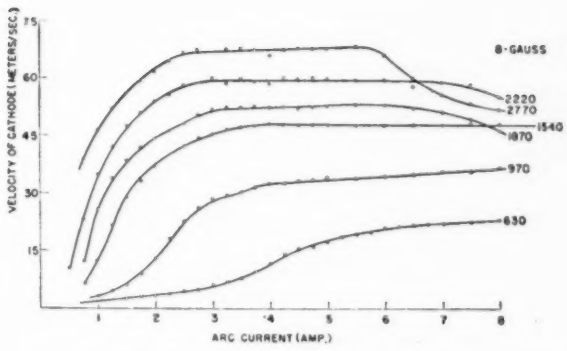


FIG. 3. Velocity of cathode spot on Hg pool *vs.* arc current for various magnetic fields. Pressure = 0.7μ.

is required to reverse the spot at high currents than at low. This suggests that the point of reversal represents a condition of balance between the forces acting on the spot and the forces acting on the column, since it has been observed that the column always tries to move in the "correct" direction.

The appearance of the rotating discharge is shown in the series of pictures in Fig. 5. These are moving pictures, taken at high speed (4000 frames per second). The arc was moving at a speed of about 75 meters per second, with a current of 3.5 amp. The column can be seen extending from the spot away from the direction of motion. The lower image is due to reflection in the magnet surface. The pictures reveal an interesting phenomenon in that occasionally there appears a large dark space between the column and the glow at the cathode. Evidence of this is beginning to appear in the second frame, with a larger gap showing in the third. The dark region has been filled with glowing gas again in the two final pictures. Examination of the complete film indicates that these gaps are randomly distributed in time with no apparent periodicity. Note that the glow of the cathode persists, regardless of the column behavior.

The phenomenon of retrograde motion and reversal was observed with many other gas and electrode combinations. Gases used included nitrogen, oxygen, hydrogen, carbon dioxide, helium, and argon, with electrodes of copper, carbon, tungsten, molybdenum, aluminum, nickel, and cadmium. There was found to be very little, if any, effect of electrode material on the reversing pressure. The principal effect of the electrode appeared to be in determining the stability of the arc. It was not possible to establish arcs at these currents between electrodes which had been carefully cleaned and outgassed. It was necessary in all cases to operate with metals which were oxidized to some extent.

The reversing pressure was sharply affected by the kind of gas used. The results obtained are shown in Fig. 6. The arc current was adjusted to 4 amp. and the magnetic field kept at a low value of about 100 gauss. There is a wide range in reversing pressures, ranging from about 1 cm for mercury to 16 cm for helium, with other gases lying in between. As Fig. 6 shows, there is some correlation between the reversing pressure and the first resonance potential of the gas. The first resonance potential may be taken as a rough measure of the voltage drop between the cathode and the column (the "cathode drop"). The first resonance potential is probably a better choice than the ionization potential because of the possibility of ionization occurring in several steps rather than a single ionizing collision. The probability of step-wise ionization is high where the current density is large. The current density at the arc cathode spot may be greater than 50,000 amp. per cm^2.[3]

It will be noted that mercury exhibits the lowest reversing pressures. For every other gas investigated, the reversing pressures were higher. In fact, retrograde motion can be observed at pressures as high as atmospheric, depending on the kind of gas, the current, and magnetic field strength. In argon with oxidized tungsten electrode, as shown in Fig. 7, the reversing pressure can be as high as 60 cm Hg, and as low as 5 cm Hg. It will be noted that the effect of current and magnetic field on the reversing pressure is, in general, the same as found in mercury, i.e., the reversing pressure is greater fields or smaller currents.

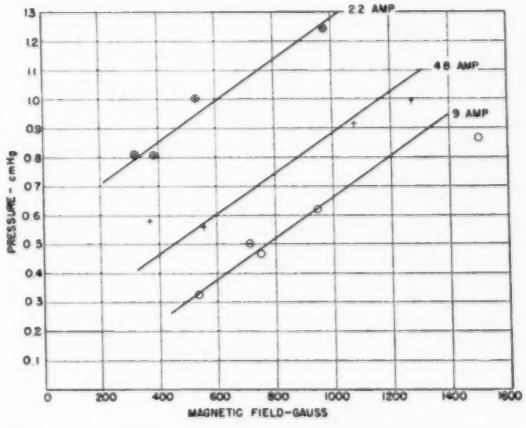


FIG. 4. Reversing pressure *vs.* magnetic field in mercury pool tube.

[3] J. D. Cobine and C. J. Gallagher, Phys. Rev. 74, 1524 (1948).

There is at present no adequate explanation for the phenomena described here. A complete theory would undoubtedly require a knowledge of the mechanism of current transfer at the arc cathode, since the action of the magnetic field must be either at the cathode surface or else at the region of high ion density immediately outside it. If one rules out the Righi-LeDuc effect, as originally proposed by Smith,[2] and other similar effects (Hall, Ettingshausen, and Nernst effects,[4] there remains only the effect of the magnetic field on electron trajectories as the primary physical phenomenon. The electrons involved are probably those emitted by the cathode. The direct effect of the magnetic field on the paths of positive ions is probably quite small, at least in comparison with the change in electron paths. However, the electron trajectories are bent in a direction opposite to the retrograde motion, so that the motion itself is a secondary effect. A simple calculation of the motion of a single electron in crossed electric and magnetic fields indicates that the particle is deflected only a small amount before colliding. This is especially true at higher pressures where the mean free path is short. Bending the electron paths would certainly shift the center of gravity of the positive ion space charge outside the cathode, since these ions are probably formed by collisions with emitted electrons. Longini[5] makes a plausible case for a net space-charge shift in the retrograde direction by considering that the positive-ion space charge is shifted less than the electron space charge. It is not clear from his picture how reversal can occur, or why the direction of motion depends on the pressure, the current, or the magnetic field. Himler and Cohn[6] have offered a less plausible explanation by considering the electrons to be emitted from the cathode in all directions, with those emitted in the "correct" direction being returned to the cathode without contributing to the current. Those emitted in the retrograde direction would produce a positive-ion space charge in that direction which would produce a field sufficient to extract more electrons from the cathode. It does not seem possible that there can be any such

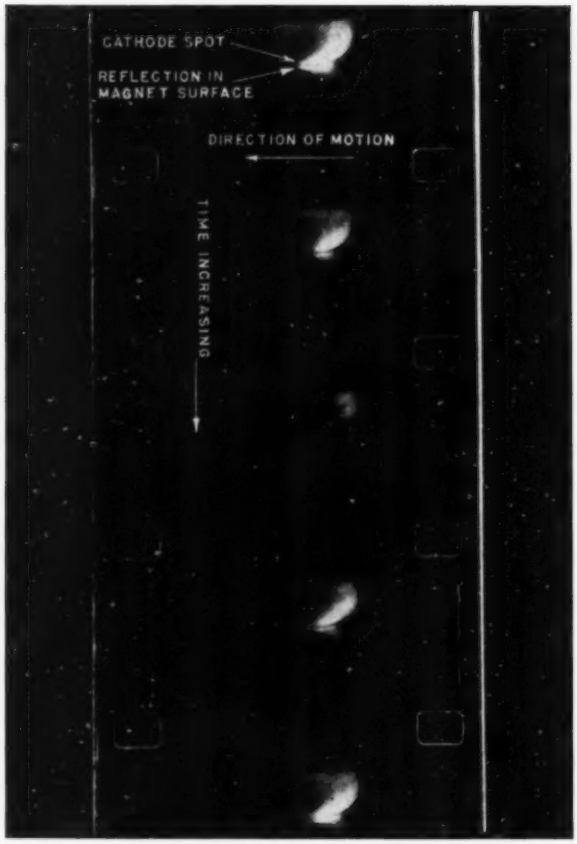


FIG. 5. High speed pictures of cathode spot moving on Hg pool (4000 frames per second).

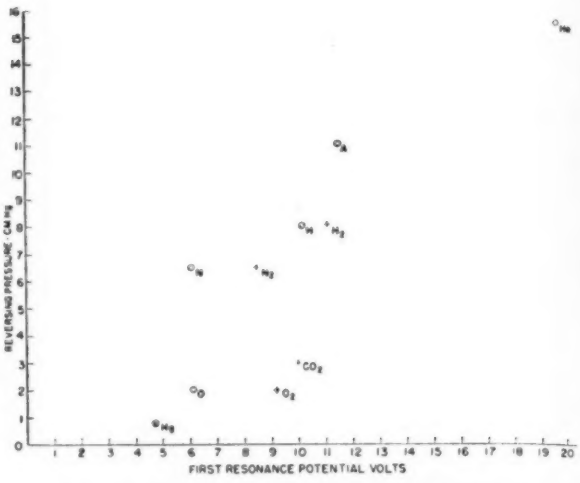


FIG. 6. Reversing pressure *vs.* resonance potential. arc current—4 amp., magnetic field—100 gauss.

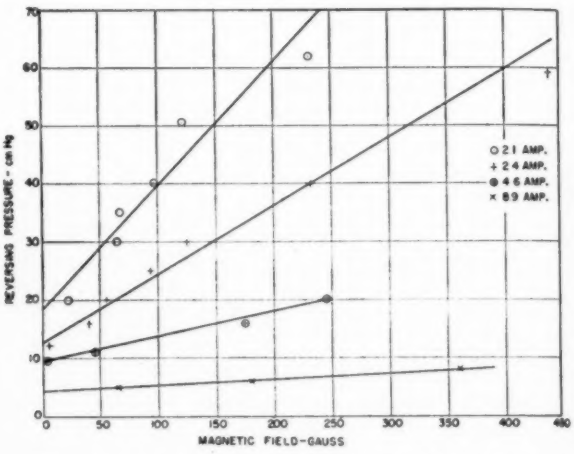


FIG. 7. Reversing pressure *vs.* magnetic field oxidized tungsten in argon.

[4] P. W. Bridgman, *The Thermodynamics of Electrical Phenomena in Metals* (The Macmillan Company, New York, 1934).

[5] R. L. Longini, Phys. Rev. **72**, 184 (1947); **71**, 642 (1947).

[6] G. J. Himler and G. I. Cohn, Elec. Eng. **67**, 1148 (December, 1948).

random distribution in directions of emission, at least, not with the spread that this picture would require.

Both of the above pictures point to the position of the positive-ion space charge as the important factor in determining the position of the cathode spot which in turn means that the emission process depends on the positive ions. That this must be so seems to be borne out by two experimental observations. First, there is the fact that the velocity of the spot is of the order of the probable thermal velocity of the gas atoms (or ions). In fact, it was not possible to get velocities quite as high as thermal velocity. Unless the electric fields are very distorted, the motions of the ions parallel to the cathode will be determined by the thermal velocities. A second evidence of the importance of the position of the positive ions is the observations by Smith,[7] and also by Cobine and Gallagher,[8] that the cathode spot does not exhibit retrograde motion if the cathode is heated to a temperature high enough to support thermionic emission. In this case, the positive-ion space charge is no longer required except to accelerate electrons which have already been emitted, so that the position of the space charge is less important in determining the positions of the cathode spot. These observations are not conclusive, but they do indicate that the positive-ion space charge plays an important role in determining the motion of the spot, and, hence, in maintaining current transfer at the cathode surface.

[7] C. G. Smith, Phys. Rev. **73**, 543 (1948).
[8] J. D. Cobine and C. J. Gallagher, Elec. Eng. **68**, 469 (1949).

Transonic Potential Flow of a Compressible Fluid*

W. R. Sears
Cornell University, Ithaca, New York
(Received February 20, 1950)

Even under the assumptions of irrotational, isentropic flow, which have been found generally useful for subsonic and supersonic cases, the equations of gas flow are relatively intractable for mixed, transonic situations. Approximate methods of solution used by a number of investigators are reviewed briefly, as well as the hodograph technique, which yields exact solutions of the equations for plane flow. It is pointed out that all the methods predict smooth, potential mixed flows involving imbedded regions of supersonic speed and both acceleration and deceleration through the speed of sound. There is no experimental verification of the existence of such flows.

Three possible explanations for this sharp discrepancy between experiment and perfect-fluid theory have been advanced; namely, (a) effects of viscosity, (b) non-existence of neighboring solutions, and (c) temporal instability. These are reviewed in turn. None has led to a complete explanation, to date. Kuo's stability calculations are described briefly. His results indicate that stable, smooth, mixed flows may exist if certain conditions are satisfied.

1. METHODS OF CALCULATING TRANSONIC POTENTIAL FLOW

MUCH of our progress in the theory of compressible fluid flows is based on the compatible assumptions of irrotational flow and isentropic changes of state. These approximations are justifiable, in many situations, because a gas of small viscosity in purely subsonic flow, or in flow involving only straight or weak shocks, does flow nearly isentropically and irrotationally. Thus, for many subsonic and supersonic flows, the results of potential-flow theories are in good agreement with experimental observations. We are concerned here with the validity of those assumptions in cases of transonic flow.

Even for steady plane flows that are inviscid, isentropic, and irrotational the differential equations that describe the flow patterns are non-linear, and therefore do not yield, directly, to standard and general mathematical methods of solution. To be sure, these equations can be made linear, by the simple approximation of neglecting the non-linear terms, and are still useful to describe situations occurring at low speeds and again at moderately high speeds, but not near the speed of sound. When the gas takes on speeds close to the sonic speed, relative to a solid obstacle or the walls of a channel, the same essential non-linearity appears in the equation of continuity that is so clearly demonstrated at the throat of a Laval nozzle. The familiar "linearized theory" then fails.

The engineer has not been completely unable to cope with this difficulty. Recognizing the essential non-linearity and the formidable mathematical difficulty of his problem, he has, with some success, taken recourse to such dependable methods as purely numerical computation and series expansion.

G. I. Taylor,[1] for example, in 1930 assumed the velocity potential to be expanded in a double power series in the Cartesian coordinates x and y,

$$\phi = a_1x + a_2x^2 + c_2y^2 + a_3x^3 + c_3xy^2 + a_4x^4 + c_4x^2y^2 + e_4y^4 + \cdots. \quad (1)$$

* An invited paper delivered at the meeting of the American Physical Society at Columbia University, February 4, 1950.

[1] G. I. Taylor, "The flow of air at high speeds past curved surfaces," Aero. Res. Comm. Reports and Memo. No. 1381, London (1930).

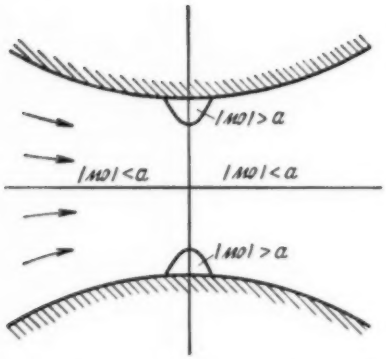


FIG. 1. Transonic potential flow through nozzle, according to calculation of Taylor (see reference 1).

He then proceeded to calculate numerical solutions to the non-linear equations, for the case of plane flow through a symmetrical nozzle. This procedure led to solutions involving both subsonic and supersonic speeds, which will be described later.

Using an entirely different approach, following Prandtl and Ackeret, a number of writers have imagined the velocity potential (or the stream function) to be expanded in powers of a thickness parameter τ, for a family of flow patterns produced around a family of slender cylindrical objects:

$$\phi = U\{x + \tau\phi_1 + \tau^2\phi_2 + \cdots\}. \qquad (2)$$

For example, the parameter τ might be the thickness ratio, for a family of symmetrical airfoils, or the camber, for a family of circular-arc profiles. This results in an iteration procedure by which the successive coefficients of τ^n are calculated from the knowledge of the preceding ones, and in fact the first approximation is precisely the result of the "linearized," or Prandtl-Glauert theory mentioned above. Although the convergence of this process has never been established theoretically, the numerical work appears to converge very satisfactorily, at least at low Mach numbers, and there is no apparent difficulty in producing flows of the transonic type; that is, having both supersonic and subsonic local speeds. The most important work along this line was done by Carl Kaplan[2] in America and by H. Görtler[3] in Germany.

Howard Emmons,[4] in 1946, undertook to attack the problem purely numerically. Approximating to the differential equations of plane motion by differences, he applied the relaxation technique successfully, making the switch from subsonic to supersonic conditions when necessary, and finally, if necessary, introducing shock waves.

More recently, Theodore von Kármán,[5] and, independently, G. Guderley,[6] discovered that a perturbation theory of a different sort, again for slender bodies, could be made for the transonic regime. In this theory it is assumed that every velocity represents a small perturbation of a uniform parallel stream at sonic speed a^*. The resulting equation for the perturbation velocity

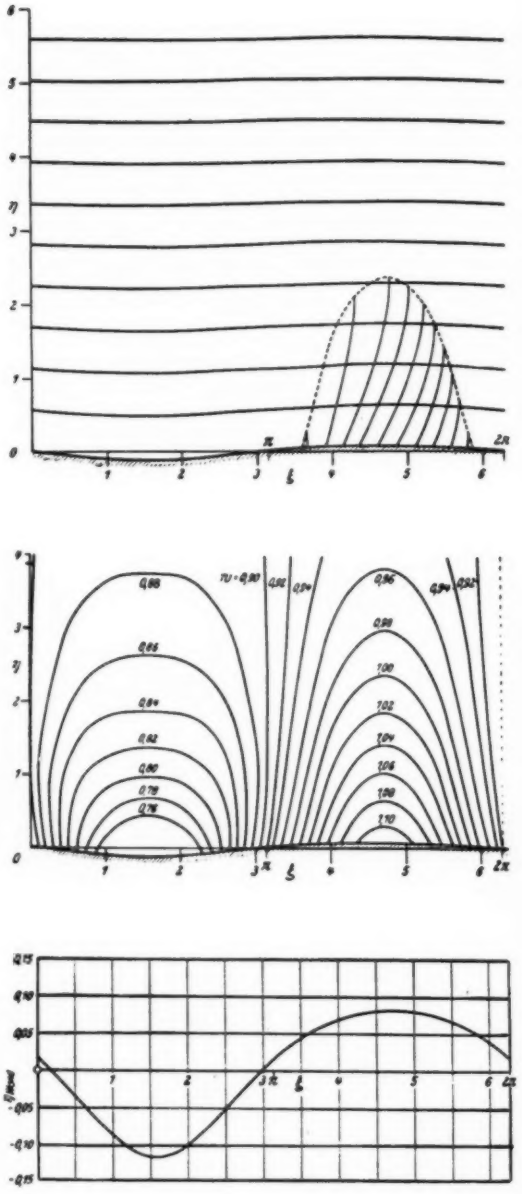


FIG. 2. Transonic potential flow over a wavy wall, according to calculations of Görtler (see reference 3). Top: streamlines and (in supersonic region) Mach lines. Center: curves of constant Mach number. Bottom: shape of the wall, ordinate enlarged ten times.

[2] Carl Kaplan, "The flow of a compressible fluid past a curved surface," N.A.C.A. Tech. Report No. 768 (1943).

[3] H. Görtler, "Gasströmungen mit Übergang von Unterschall- zu Überschallgeschwindigkeiten," Zeits. f. Angew. Mat. Mech. **20**, 254–262 (1940).

[4] H. W. Emmons, "The theoretical flow of a frictionless, adiabatic, perfect gas inside of a two-dimensional hyperbolic nozzle," N.A.C.A. T.N. 1003 (1946). Also, "Flow of a compressible fluid past a symmetrical airfoil in a wind tunnel and in free air," N.A.C.A. T.N. 1746 (1948).

[5] Th. von Kármán, "The similarity law of transonic flow," J. Math. and Phys. **26**, 182–190 (1947).

[6] G. Guderley, "On the transition from a transonic potential flow to a flow with shocks," U.S.A.F. A.M.C. Tech. Report No. F-TR-2160-ND (1947).

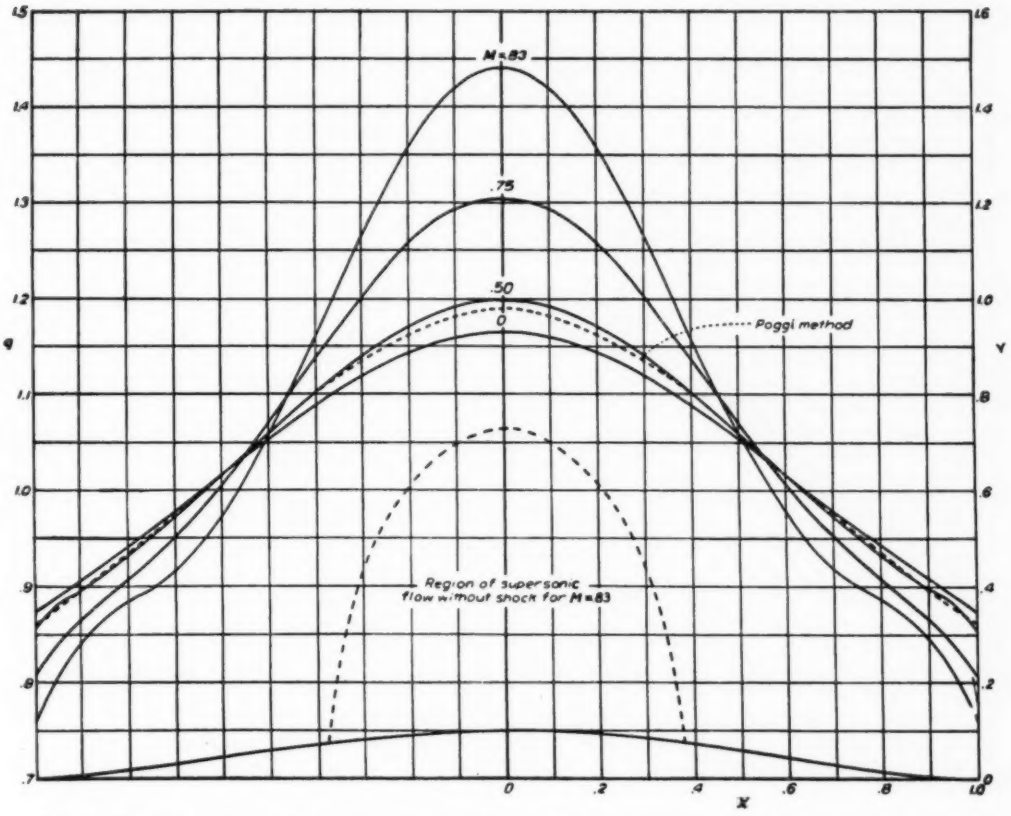


FIG. 3. Transonic potential flow over a bump, according to calculations of Kaplan (see reference 2). Velocity distribution on the surface for several stream Mach numbers M, and extent of supersonic flow for $M=0.83$.

potential $\phi(x, y)$ is

$$\frac{\partial^2\phi}{\partial y^2}=\frac{\gamma+1}{a^*}\frac{\partial\phi}{\partial x}\frac{\partial^2\phi}{\partial x^2}, \tag{3}$$

where γ denotes the isentropic exponent. This is typically non-linear and exhibits the characteristic change-over from elliptic to hyperbolic type at the speed of sound. The discovery that underlies this theory is that, for a slender body of thickness ratio τ, the partial derivatives have the following relative orders of magnitude:

$$\left.\begin{aligned}&\partial(\ )/\partial x=O(1),\quad \partial(\ )/\partial y=O(\tau^{\frac{1}{3}}),\\ \text{while}\qquad &\phi=O(\tau^{\frac{2}{3}}).\end{aligned}\right\} \tag{4}$$

Moreover, both von Kármán and Guderley pointed out the similarity rule that is implied by these results, and which gives the engineer badly needed assistance in filling in the gap in his knowledge around the speed of sound. At the same time, Eq. (3) gives the theoretician a relatively simple equation with which to study the essential features of transonic flow. Certainly its potentialities have hardly been exploited to date.

Finally, there is one more device for the study of these inherently non-linear flow equations, at least in two-dimensional cases; and that is to transform them to the hodograph plane, i.e., to introduce a change of variables by which the flow velocity components u and v, say, become the independent variables, rather than x and y. The equations then become linear, which is not completely surprising because the essential non-linearity arose from the fact of the density's being a function of the speed, $(u^2+v^2)^{\frac{1}{2}}$. This method of linearization was discovered by Molenbroek[7] in 1890, and was applied to this problem in 1904 by the Russian, Chaplygin,[8] working on the theory of gas jets. It permits the construction of solutions by superposition of elementary solutions—in this case certain hypergeometric functions. The essential difficulty, unfortunately, then resides in the question of satisfying boundary conditions. Naturally, one does not in most cases have any way to satisfy the conditions for flow around a given solid contour in terms of boundary conditions established in the hodograph plane.

Nevertheless, the method has the unique advantage of supplying exact solutions of the equations of steady plane potential flow of a perfect gas. Thus its importance

[7] P. Molenbroek, "Über einige Bewegungen eines Gases bei Annahme eines Geschwindigkeitspotentials," Archiv d. Mat. Phys. **9**, 157–195 (1890).

[8] S. A. Chaplygin, "On gas jets," Scientific Annals of the Imperial University of Moscow, Physico-mathematical Division, No. 21, Moscow, 1904 (translated in N.A.C.A. Tech. Memo. No. 1063 (1944)).

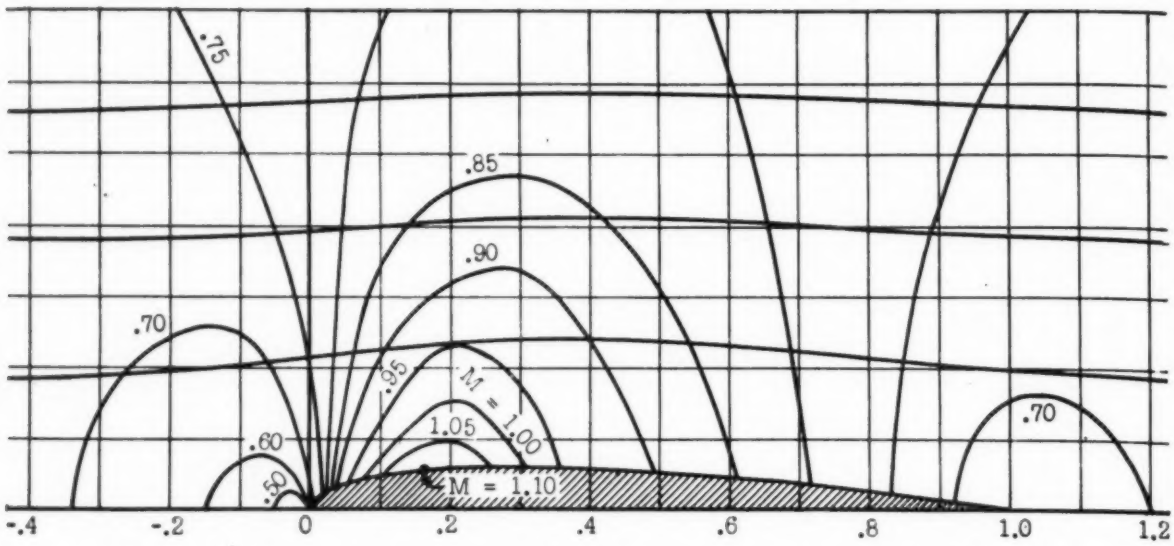


FIG. 4. Transonic potential flow over an airfoil, according to calculations of Emmons (see reference 4). Streamlines and lines of constant Mach number.

in research, and as a test for approximate methods, is great. Being practically unable to satisfy the boundary conditions of given shapes, one naturally tries the next best thing, which is to find a likely solution, either by trial or by relating it to incompressible flow in the proper limiting case, and then to see what sort of physical situation it gives. There are a number of different mathematical techniques for carrying out this sort of calculation, and unfortunately they all seem to be pretty complicated. The hodograph transformation is practically always multiple-sheeted with branch points that cause great difficulties in series expansions. Consequently, in many of these investigations the physics seems to be buried under tremendous amounts of analysis, and there is also an apparently endless controversy, disconcerting to the outsider, among various students of the subject, who fail to find much of value in each other's methods.

2. THEORETICAL MIXED FLOWS

Nevertheless, the hodograph method, in common with all the other, approximate, methods that have been applied to the transonic flow problem, has yielded one result of startling significance to the aeronautical engineer. It produces isentropic, potential flows that involve local acceleration through the velocity of sound, an imbedded region of supersonic flow near a body, and deceleration back to subsonic speeds. This is intriguing because it contrasts so radically with the situation commonly observed in nature—where the transition from supersonic to subsonic flow is accomplished by means of a violent, non-isentropic process, namely, a shock wave.

Apparently the first example of this theoretical phenomenon was provided by Taylor,[1] who found it in a symmetrical nozzle, using the series expansion mentioned above. One of his results is reproduced in Fig. 1. Analogous results were obtained by Görtler[3] (Fig. 2) and Kaplan[2] (Fig. 3), both using the method of expansion in powers of the thickness, and by Emmons[4] in his purely numerical solutions (Fig. 4).

The first exact solution of this type was found by Ringleb,[9] by means of the hodograph transformation. Ringleb started with a very simple solution of the hodograph equations, being, in fact an almost obvious generalization of the incompressible case of flow around a 180° corner. His result is reproduced in Fig. 5. It is not only of interest as the first exact mixed solution, but it presents two other surprising features. The first of these is the great magnitude of the supersonic Mach numbers attained in the flow (about 2.5), and the second is the appearance of singularities, called "limit lines" which seem to invalidate the solution when too great wall curvature is attempted. In other words, any two streamlines such as (a) and (b) in Fig. 5 might form the walls of a channel, and according to the theory, this channel would permit isentropic acceleration to, and deceleration from, supersonic speeds. But streamlines between (c) and (d) cannot be traced beyond the limit line; according to the transformation, in fact, they double back on themselves there and cross other streamlines—in other words, they go over onto a different Riemann sheet.

When Ringleb's paper first appeared, in 1940, there was considerable speculation about the physical meaning of the limit line, and there was a natural tendency to identify them with shock waves. More recently, however, aerodynamicists have taken a more cautious

[9] F. Ringleb, "Exakte Lösungen der Differentialgleichungen einer adiabatischen Gasströmung," Zeits. f. angew. Mat. Mech. **20**, 185–198 (1940). Abstract in J. Roy. Aero. Soc. **46**, 403–404 (1942).

position on this question—although it remains controversial. In Ringleb's case, for example, the appearance of the limit line at streamline (c) is accompanied by infinite curvature of that streamline. There is considerable doubt as to whether it could ever appear, anywhere in the flow, in the case of flow over a smooth solid body of fixed shape. This question has been studied by Tollmien,[10] Schaefer,[11] Nikolskii and Taganov,[12] and Friedrichs,[13] not all of whom have arrived at the same conclusions. However, it is not the principal concern of this paper, and we shall leave it at this point.

The resemblance of Ringleb's flow pattern to the flow over the top of an airfoil at high subsonic speed inspired Tsien and Kuo[14] to attack problems of this type with the hodograph method; i.e., flow of a parallel, subsonic stream past a cylindrical obstacle. One of their results is shown in Fig. 6. Here we have a symmetrical, nearly elliptic cylinder, in plane potential flow, and the Mach numbers in its imbedded supersonic zone reach a maximum value of 1.24.

It should be clear at this point that perfect-fluid theories, both exact and approximate, predict smooth potential mixed flows, in channels and past obstacles, that involve both acceleration and deceleration through the speed of sound. It is obviously of great practical and scientific importance to explore thoroughly the prospects of achieving this type of flow experimentally. The Tsien-Kuo cylinder clearly produces no drag, in the absence of skin friction, while the shock wave, if one occurred at a Mach number near 1.24, would be of strength sufficient to involve considerable drag, and moreover would probably cause boundary-layer separation and all of the undesirable consequences thereof. There would be great merit in learning how to reproduce this smooth mixed flow in nature.

The absence, to date, of any convincing experimental confirmation of this type of flow, however, is impressive. Invariably, when the local supersonic speeds of a transonic situation are great enough to make the determination reliable, we find that the supersonic region terminates in a shock wave, or in a series of shock waves, rather than in a smooth deceleration through the speed of sound.

3. POSSIBLE EXPLANATIONS

What is the reason for this glaring discrepancy between theory and experiment? Three different plausible explanations have been suggested: (a) that the theory is in error by the neglection of viscosity; (b) that the potential flows exist for only special shapes and flow speeds and do not have neighboring solutions. Hence they cannot be actually produced in experiments; (c) that the potential flows are unstable to small disturbances in the airstream.

In support of the first of these, Professor Liepmann has pointed out[15] that the smooth deceleration can be thought of as the result of effects carried rearward by the family of Mach waves originating farther forward, as sketched in Fig. 7(a), and that a distortion of the contour, as might be caused by a boundary layer, would suffice to cause an envelope of waves, as in Fig. 7(b). Such a discontinuity, of course, marks the breakdown of potential flow, and the pattern must be replaced by one involving a shock. But this would imply that a slightly different body, producing the Mach waves of Fig. 7(a) outside of its boundary layer, would permit smooth

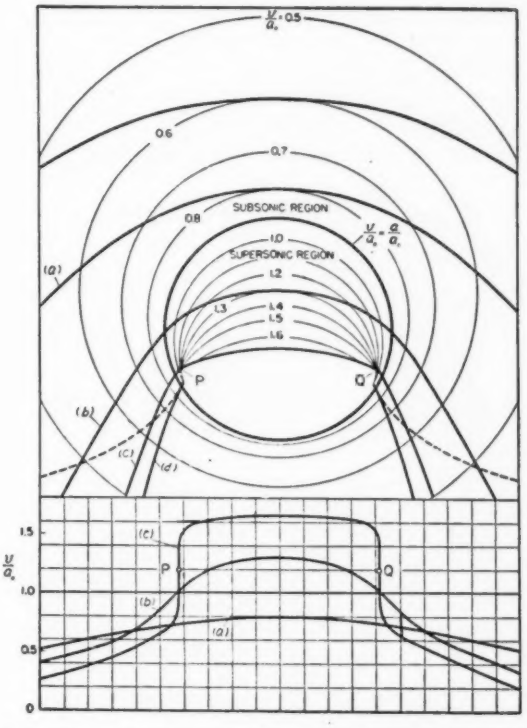


FIG. 5. Transonic flow around an obstacle, according to calculations of Ringleb (see reference 9). Top: streamlines and lines of constant speed. Bottom: velocity distributions along three streamlines.

[10] W. Tollmien, Grenzlinien adiabatischer Potentialströmungen. Zeits. f. angew. Mat. Mech. **21**, 140–152 (1941).

[11] M. Schaefer, Enveloppenbildung krummliniger Machscher Wellen bei Strömung längs einer konvexen Wand, Technische Hochschule Dresden, 1944 [translated in U.S.A.F. A.M.C. Tech. Report No. F-TS-1203-1A (GDAM A9-T-13)]. Also, Auftreten eines Verdichtungsstosses in der Nachbarschaft einer konvexen singularitätenfreien Wandkontur. Technische Hochschule Dresden, 1944 [translated in U.S.A.F. A.M.C. Tech. Report No. F-TS-1206-1A (GDAM A9-T-14)].

[12] A. A. Nikolskii and G. I. Taganov, "Gas motion in a local supersonic region and conditions of potential-flow breakdown," Prikladnaya Matematika i Mekhanika **10**, 481–502 (1946). [Translated in N.A.C.A. Tech. Memo. No. 1213 (1949).]

[13] K. O. Friedrichs, "On the non-occurrence of a limiting line in transonic flow," Inst. for Math. and Mech., New York University, Rep. IMM-NYU 165 (1947).

[14] H. S. Tsien and Y. H. Kuo, "Two-dimensional irrotational mixed subsonic and supersonic flow of a compressible fluid and the upper critical Mach number," N.A.C.A. T.N. 995 (1946). Also, Y. H. Kuo, "Two-dimensional irrotational transonic flows of a compressible fluid," N.A.C.A. T.N. 1445 (1948).

[15] H. W. Liepmann, "Investigations of the interaction of boundary layer and shock waves in transonic flow," Final Report, Contract W33-038 ac-1717 (11592), supplemental agreement No. 4 (S-4843), U.S.A.F. A.M.C., Wright Field, Dayton, Ohio.

potential flow. This proposal hardly seems adequate to explain the complete absence of flows of this type in experiments.

Moreover, if Liepmann's explanation is correct, it should be possible to achieve shockless mixed flows experimentally with the aid of boundary-layer removal. It seems important to decide whether or not the inviscid potential flow itself has any peculiarities that would make it impossible.

The second proposal, that the flows are very special ones, with no neighboring flows, is supported by Guderley and Busemann. Their physical argument is illustrated by Fig. 8, which is taken from Busemann's recent paper.[16] Suppose there exists a smooth mixed flow, and that a slight change is made in the body contour as shown. Since the local flow is supersonic, the effects of this alteration, in the new steady-flow pattern, will be confined to an "alley" leading to the sonic curve, from which point it is reflected back to the body, and so forth, as sketched. But according to this process it finally affects the nearly-sonic flow toward the rear, and we know that the pressure, density, and speed increments will be magnified in such a region. Thus a small change of the boundary shape must produce a large change in the flow pattern; i.e., there is no neighboring solution. Substantially the same argument has been put forth by Frankl.[17]

All this is quite analogous to the case of decelerating channel flow. It is easy to draw up a channel for smooth deceleration of supersonic flow, through sonic, to subsonic; in fact, the channel is simply a Laval nozzle in reverse (Fig. 9). But the flow is unreal, because there is no neighboring solution. A slight change of the contour at A produces violent effects at B. The real flow is the one involving a shock wave.

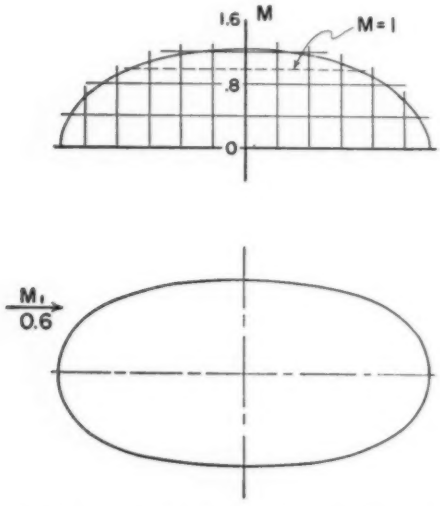


FIG. 6. Mach number distribution along a cylinder in transonic potential flow at a stream Mach number of 0.6, according to calculations of Tsien and Kuo (see reference 14).

Thus, Busemann's is certainly a very plausible explanation. It would be indisputable if the perturbation at the sonic curve did not actually affect the whole flow pattern through the subsonic flow, but of course it does. To carry along the channel analogy, the trouble at B (Fig. 9) can be avoided by a small alteration at C or C', or by an alteration upstream, as at D or D', which would cancel the effect at A. Busemann argues that in the case of flow past a body it is unlikely that the effect on the subsonic flow will be just what is required to neutralize the disturbance and prevent the large changes toward the rear. Perhaps it does seem unlikely, but what is unlikely is often true.

Guderley[6] has attempted to give a mathematical theory of this effect, showing that there are eigensolutions that would invariably be brought in by any alteration of the contour, and whose coefficients increase without limit as the sonic line is approached. Unfortunately, Guderley's calculation becomes so complicated, requiring further approximations and assumptions as it progresses, that its conclusions are not convincing. For example, these solutions have infinite amplitudes at both the upstream and downstream sonic points, and since we postulate smooth flow upstream, it is not clear why they should not all be rejected, removing the trouble at the downstream end as well. I suspect that the singularities are not real ones, but have arisen from Guderley's method of approximation, and the form of his results seems to confirm this.

There is also the problem of reconciling the Guderley-Busemann-Frankl theory with the Taylor, Emmons, and Kaplan cases mentioned above. None of these were calculated for a peculiar or special body shape or for any unique combination of body shape and Mach number. It seems preposterous to think that they are some kind of singular solutions, without neighbors. But these are all approximate solutions, and how much can one conclude about existence and uniqueness from investigations of this sort? Again it seems "unlikely" that these calculations would not show some trouble if they were being forced to fit contours not specially suited for this type of flow. For example, one might have thought that Emmons' numerical work would diverge. This point is not conclusive, however, for the effects may be small at the Mach numbers handled, and may not have been noticed. Moreover, returning to our channel analogy, we must admit that there an approximate method (one-dimensional gas dynamics) fails entirely to predict the non-existence of the flow of Fig. 9.

4. THE STABILITY OF MIXED FLOWS

In this situation, my colleague, Professor Kuo,[18] has undertaken to investigate the stability of mixed po-

[16] A. Busemann, "The drag problem at high subsonic speeds," J. Ae. Sci. 16, 337–344 (1949).

[17] F. I. Frankl, "On the formation of shock waves in subsonic flows with local supersonic velocities." Prikladnaya Matematica i Mekhanika 11, 199–202 (1947). (Translated in USAF–AMC Tech. Report No. F-TS-1223-IA, Dayton, Ohio, May, 1949.)

[18] Y. H. Kuo, "On the stability of two-dimensional smooth transonic flows," paper delivered at the Annual Meeting of the

tential flow by the standard method of tracing the histories of time-dependent small disturbances. To a considerable extent his method follows that of Kantrowitz,[19] who carried out an approximate, but highly illuminating investigation of the stability of decelerating channel flow. Kantrowitz considered a channel flow going from supersonic to subsonic and calculated the behaviors of expansion and compression disturbances coming from downstream. He pointed out that the compression part of either a compression or expansion pulse steepens to form a weak shock (as Riemann showed for a uniform field) and the pulse assumes a characteristic triangular shape. Furthermore, he showed that the expansion pulses are unable to penetrate the supersonic region, and, in fact, are "consumed" by their own shock waves. Compression pulses, on the other hand, which have shock waves at their heads, move into the supersonic flow and pile up there, so that the flow finally undergoes violent changes.

Following Kantrowitz, Kuo considers the effects of upstream-moving disturbances of the compression-pulse type, and assumes that they have undergone the Riemann steepening process and have taken on the characteristic triangular shape. Since the sonic region is most critical, and since disturbances travel slowly in this region, he assumes that the pulses take on the character typical of sonic flow; i.e., their derivatives have the relative magnitudes specified by the Kármán-Guderley transonic similarity theory and stated in Eqs. (4). This leads to an accompanying assumption about the relative size of the time derivatives; it turns out that there is only one possibility here that gives a reasonable result for the non-stationary flow.

These assumptions are sufficient to permit a solution for small disturbances superimposed on a steady potential mixed flow of a parallel stream past a cylinder, at least for the region near either sonic boundary.

Let the velocity potential of the flow be

$$\varphi'(x, y, t)=\Phi(x, y)+\phi(x, y, t), \tag{5}$$

where $\phi(x, y, t)$ denotes the potential of the small disturbance superimposed on the steady transonic flow $\Phi(x, y)$. Then, by virtue of the order-of-magnitude argument for the sonic region, the differential equation for $\phi(x, y, t)$ is reduced to the following approximate one:

$$(\gamma+1)(\Phi_x^{(1)}+\phi_x)\phi_{xx}-a^*\phi_{yy} +2\phi_{xt}+(\gamma+1)\Phi_{xx}^{(1)}\phi_x=0, \tag{6}$$

where $\Phi^{(1)}$ denotes $\Phi-a^*x$. Kuo approximates to $\Phi^{(1)}$ in linear form; i.e.,

$$\Phi^{(1)}=\alpha\xi+\beta\eta, \tag{7}$$

where α and β are constants, while ξ and η denote

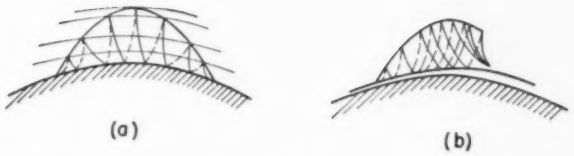


FIG. 7. Sketch showing effect of boundary layer on transonic potential flow: (a) without viscosity; (b) with viscosity.

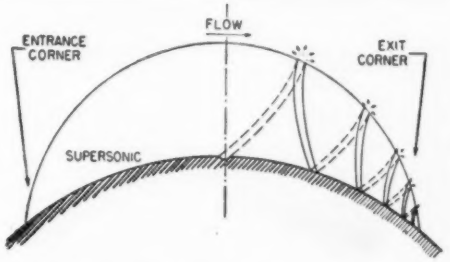


FIG. 8. Sketch showing effect on transonic potential flow of small alteration of boundary shape, according to Busemann (see reference 16).

distances measured parallel and perpendicular to the body contour from the sonic point. The boundary condition at the solid body is, similarly,

$$\phi_\eta=\phi_\xi(m+k\xi), \tag{8}$$

where m is the slope (relative to the free-stream direction) and k the curvature of the body surface at the sonic point.

Kuo finds the solution $\phi(x, y, t)$ in simple functions involving, besides initial conditions, the parameters

$$\left.\begin{aligned}\nu&=(\gamma+1)\alpha\\ \text{and}\qquad\sigma&=\tfrac{1}{2}(\nu+k/m).\end{aligned}\right\} \tag{9}$$

In the case of one-dimensional flow, his results reduce to those of Kantrowitz.[19]

For convex surfaces, $k<0$. At the accelerating sonic point, $\alpha>0$ and, presumably, $m>0$; this combination makes all the disturbances damp rapidly to zero, which confirms the observed stability of such flow. On the other hand, where $k<0$ and $\alpha<0$ (decelerating steady flow over a convex surface), the damping or growing of triangular pulses is influenced by the sign of m. When the decelerating sonic point occurs in a region of negative slope m, as is the case in Figs. 1–6, the disturbances increase in size in the sonic region, which indicates instability. But the theory brings forth an intriguing new idea: if m were positive here, the flow might again be stable. For example, if the imbedded supersonic area near an airfoil could be made to lie entirely *forward* of the maximum ordinate of the surface (Fig. 10), there would be a great reduction in the rate of growth of pulses as they pass through the sonic area, so that much larger disturbances could be tolerated; this might be termed pseudostability. Moreover, certain combinations of α, β,

Institute of Aero. Sciences, New York, January 24, 1950 (to be published).

[19] A. R. Kantrowitz, "The formation and stability of normal shock waves in channel flows," N.A.C.A. T.N. 1225 (1948).

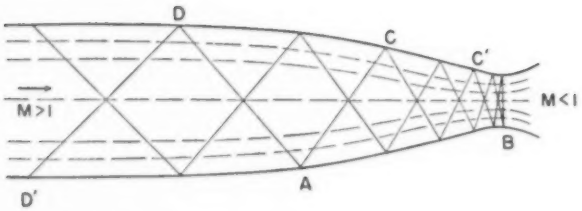


FIG. 9. Sketch showing reversed Laval nozzle for deceleration of supersonic flow to subsonic.

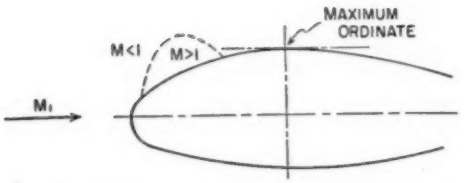


FIG. 10. Sketch showing possible stable transonic potential flow according to Kuo.

and m would result in complete damping of the pulses, i.e., stability.

So far as we know, there are no definitive measurements for flows of this type. Apparently it is possible to find airfoil shapes that have their supersonic regions entirely forward of their maximum ordinates. We are not yet sure that the peculiar combinations of α, β, and m mentioned above, for which the pulses are completely damped out, do occur. As Kuo has pointed out, there is a remarkable coincidence between his theoretical prediction and some recent experimental results of the N.A.C.A.[20] Entirely independently of Kuo's work, Nitzberg and Crandall at Ames Laboratory arrived at the empirical conclusion that the relative location of the sonic area and the "crest" of the airfoil is of significance in predicting the drag-divergence Mach number, M_D. To attribute this result to Kuo's stability effects is a long step, implying as it would that a shockless potential mixed flow exists before drag divergence or at least that the growth of small disturbances calculated by Kuo is the cause of drag divergence. A more conservative explanation would be that, although there are stationary shock waves, rather than potential flow, both below and above M_D, the shock waves do not cause separation of the boundary layer until $M=M_D$. Nevertheless, the coincidence is striking and, as Kuo has stated, seems to justify further research.

[20] G. E. Nitzberg and S. Crandall, "A study of flow changes associated with airfoil section drag rise at supercritical speeds," N.A.C.A. T.N. 1813 (1949).

5. CONCLUSION

In conclusion, it may be stated with assurance, that progress is now being made on this important question of shockless transonic flows. It might be pointed out explicitly that temporal instability and non-existence of neighboring solutions may represent the same phenomenon. One can explain the instability of an Euler column, for example, in either way, and Kantrowitz's results show that this is also true for the decelerating channel. Nevertheless, Guderley's, Frankl's, and Busemann's arguments have not suggested any stable decelerating cases, as Kuo's criterion suggests.

It seems advisable to undertake experimental confirmation of Kuo's results as soon as possible. We are studying this possibility at Cornell, but are realizing that the experimental difficulties may be considerable. One would like to have a really smooth potential flow to begin with, and to watch preselected disturbances travel into it. Unfortunately, there are usually an alarming number of disturbances naturally present, apparently originating in the boundary layer. We should like to suppress these, or to have so stable a flow that a disturbance larger than these would still be damped out. So far, we have not figured out how to accomplish this.

Analysis of Intensities Obtainable Using Pinhole and Slit Collimators in X-Ray Powder Diffraction Techniques

LEROY ALEXANDER
Department of Research in Chemical Physics, Mellon Institute, Pittsburgh, Pennsylvania
(Received February 6, 1950)

Other factors being equal, the integrated intensity of a powder diffraction reflection is a function of the geometry of the source, collimator, and sample. By means of a geometrical and algebraic analysis one can predict the relative intensities to be expected with pinhole or slit collimators of different dimensions. Important gains in intensity can be obtained by replacing pinholes with slits, and the advantage is about the same if the x-ray focal spot is viewed either longitudinally or laterally through slits of sufficient length. These results confirm experimental results reported by other investigators and also re-emphasize the potential value of the "side" windows of x-ray tubes, which are commonly neglected.

SEVERAL investigators[1–3] have recently pointed out the practicability of preparing Debye-Scherrer x-ray diffraction photographs by employing a pair of parallel slits rather than pinholes. The gain in intensity is great, being roughly proportional to the ratio of the slit length to the diameter of the pinholes employed in customary practice (for the case of the pinhole diameter equal to the slit width). The increased intensity is gained at the expense of decreased resolution and a shift in line positions, but both of these effects are very small for moderate slit lengths,[3] whereas the attendant intensity gain is several fold. Consequently such an arrangement is entirely satisfactory for the numerous applications of powder diffraction for which optimum line quality is not a prime consideration, and furthermore it permits a very substantial saving in exposure time and increase of x-ray tube life.

If the focal spot is viewed at some fixed small angle to the target face, say 6° or 7°, and if the two slits, the long dimension (if any) of the projected focal spot, and the sample axis are all parallel, it has been found[1,3] that for slits of sufficient length substantially equal gains in intensity are obtained whether the source is viewed (a) in a direction parallel to its long axis so that it appears foreshortened, the usual orientation when using pinhole techniques, or (b) in a direction normal to its long axis. In the present discussion orientations (a) and (b) will be respectively denoted the longitudinal and lateral arrangements. Thus by a relatively simple modification in camera design the "side" windows of an x-ray tube, commonly neglected by most diffractionists, can be put to good use. For this reason the recommendation has recently been made that x-ray tube manufacturers build four- rather than two-window diffraction tubes.[2]

The relative intensities to be expected with different collimator and source arrangements can be predicted with considerable accuracy by a geometrical and algebraic analysis. The mathematical development will be simplified by making the following initial assumptions: (a) The x-ray focal spot is a rectangle of uniform intensity. (b) The source-to-sample distance, x_3, is kept constant. If it were to be varied, an additional factor $1/x_3^2$ would need to be applied to each intensity equation. (c) The focal spot is viewed at some constant angle, α, with the target face. This requirement must be met since at angles less than about 8 or 10° the energy flux from the focus diminishes appreciably with decreasing angle.[4] (d) The x-ray beam impinges perpendicularly upon the sample, which takes the form of a flat sheet of negligible absorbing power for the x-rays. In actuality the results apply without serious error to a cylindrical Debye-Scherrer sample of low or moderate absorbing power. (e) The apertures S_1 and S_2 are rectangular or square in shape.

Let us set up a three-dimensional rectangular coordinate system with origin at the center of the x-ray tube focus as shown in Fig. 1. It will be noted that the widths of the slits S_1 and S_2 and the projected dimension of the focal spot as governed by the angle of view, α, are measured in the direction Z. The distances separating the source, slits, and sample are measured in the direction X; and the heights of the slits, the height of the area of sample irradiated, and the *non-projected* dimension of the x-ray focus are measured in the direction Y. For reasons that will become apparent as the analysis proceeds, one may consider the effects of the Z dimension as being virtually constant for all experimental

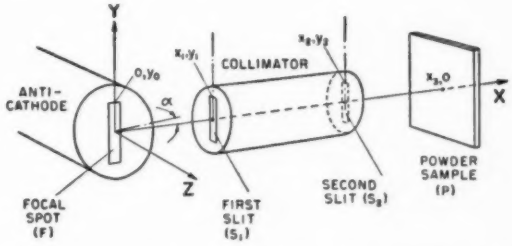


FIG. 1. Geometrical features affecting the intensity of irradiation of a powder diffraction sample. The case portrayed is that in which the focal spot is viewed laterally.

[1] A. Unmack and A. T. Jensen, J. Sci. Inst. Phys. Ind. **25**, 399 (1948).
[2] A. Taylor, J. Sci. Inst. Phys. Ind. **26**, 61 (1949).
[3] R. I. Garrod, J. Sci. Inst. Phys. Ind. **26**, 162 (1949).
[4] A. Bouwers and P. Diepenhorst, Fortschr. a. d. Geb. d. Röntgenstr. **38**, 894–898 (1928).

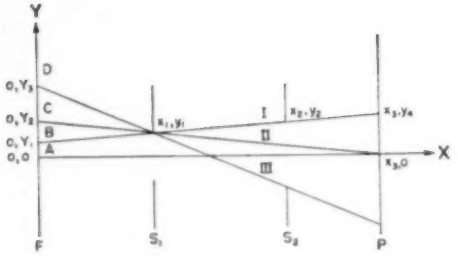


FIG. 2.

cases in which the apertures S_1 and S_2 have the same width (Z dimension) and in which at the same time the aperture width is either roughly equal to or larger than the projected (Z) dimension of the focal spot. These conditions obtain in the usual experimental cases, such as those to be compared in the present communication, and hence we may simplify the mathematical treatment by confining it to the XY plane.

Referring again to Fig. 1, the coordinates of the relevant features of our apparatus are: ends of focus, $(0, y_0)$ and $(0, -y_0)$; ends of slits 1 and 2, (x_1, y_1), $(x_1, -y_1)$ and (x_2, y_2), $(x_2, -y_2)$ respectively; intersection of x-ray beam axis and sample trace, $(x_3, 0)$. Four principal mathematical cases arise, depending upon the relative values of x_1, x_2, x_3, y_1, y_2, and y_0. Referring to Fig. 2, we may characterize these cases as A, B, C, and D according to whether y_0 falls in region A, B, C, or D respectively. Let Y_1, Y_2, and Y_3 be the respective zone boundaries AB, BC, and CD. These zone boundaries are the Y intercepts of the lines I, II, and III, and by consideration of the diagram their equations are seen to be

$$Y_1=y_1-\left(\frac{y_2-y_1}{x_2-x_1}\right)x_1=\frac{x_2y_1-x_1y_2}{x_2-x_1}, \tag{1}$$

$$Y_2=y_1+\left(\frac{y_1}{x_3-x_1}\right)x_1=\frac{x_3y_1}{x_3-x_1}, \tag{2}$$

$$Y_3=y_1+\left(\frac{y_1+y_2}{x_2-x_1}\right)x_1=\frac{x_1y_2+x_2y_1}{x_2-x_1}. \tag{3}$$

The four principal mathematical cases may then be expressed as follows:

Case A. $Y_1\geq y_0$.
Case B. $Y_2\geq y_0\geq Y_1$.
Case C. $Y_3\geq y_0\geq Y_2$.
Case D. $y_0\geq Y_3$.

Further geometrical cases arise when the apertures S_1 and S_2 differ very greatly in length. Thus when $y_1\ll y_2$, Y_1 becomes negative, and when $y_1\gg y_2$, y_4 becomes negative (see Fig. 3). However, these unusual arrangements have little practical value, and it is desirable to avoid complicating the present discussion by introducing these additional geometrical cases.

Before carrying out the mathematical analyses it will be useful to write the equations for the intersections of five lines with the line x_3 (see Fig. 3):

$$y_3=y_0+(y_1-y_0)\frac{x_3}{x_1}=y_1+(y_1-y_0)\left(\frac{x_3-x_1}{x_1}\right), \tag{4}$$

$$y_4=y_1+(y_2-y_1)\left(\frac{x_3-x_1}{x_2-x_1}\right)$$

$$=y_2+(y_2-y_1)\left(\frac{x_3-x_2}{x_2-x_1}\right), \tag{5}$$

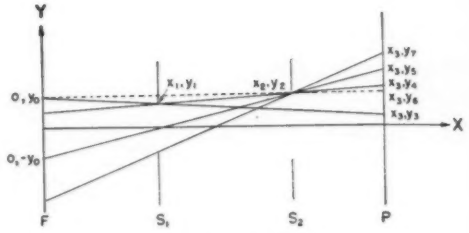


FIG. 3.

$$y_5=-y_0+(y_0+y_2)\frac{x_3}{x_2}=y_2+(y_0+y_2)\left(\frac{x_3-x_2}{x_2}\right), \tag{6}$$

$$y_6=y_0-(y_0-y_2)\frac{x_3}{x_2}=y_2-(y_0-y_2)\left(\frac{x_3-x_2}{x_2}\right), \tag{7}$$

$$y_7=-y_1+(y_1+y_2)\left(\frac{x_3-x_1}{x_2-x_1}\right)$$

$$=y_2+(y_1+y_2)\left(\frac{x_3-x_2}{x_2-x_1}\right). \tag{8}$$

We now proceed to analyze the rather typical case B. Figure 4 shows the geometrical properties of this case. At the right of the sample trace, P, the intensity of illumination of the sample as a function of y is depicted by the hatched contour. The relative integrated intensity of illumination may in general be represented by

$$I=\int_{-\infty}^{+\infty} I_y dy,$$

which becomes because of symmetry

$$I=2\int_0^{\infty} I_y dy. \tag{9}$$

For the present case the function I_y is different for each of the regions $0-y_3$, y_3-y_4, and y_4-y_5, so the integral for convenience may be broken up into three terms:

$$I=2\int_0^{y_3} I_y dy+2\int_{y_3}^{y_4} I_y' dy+2\int_{y_4}^{y_5} I_y'' dy.$$

These three integrals are seen by inspection of Fig. 4 to be given respectively by the areas of a rectangle, trapezoid, and triangle, so that the expression for I becomes

$$I=2[y_3I_3+\tfrac{1}{2}(y_4-y_3)(I_3+I_4)+\tfrac{1}{2}(y_5-y_4)I_4].$$

At any point in the zone $-y_3$ to $+y_3$ the entire focal spot can be seen and, hence, the intensity of irradiation is a maximum, I_m. Substituting I_m for I_3 and combining terms, we obtain

$$I=I_m(y_3+y_4)-I_4(y_3-y_5) \quad \text{(case B)} \qquad (10)$$

in which y_3, y_4, and y_5 are given by Eqs. (4), (5), and (6) respectively. Now the intensity I_4 is equal to I_m times the fraction of the focal spot visible at the point x_3, y_4. By comparing Figs. 2 and 4 this is seen to give

$$I_4=I_m\cdot\frac{y_0+Y_1}{2y_0},$$

$$=(1/2)I_m\left[\frac{x_2y_1-x_1y_2}{y_0(x_2-x_1)}+1\right]. \qquad (11)$$

Substitution of Eqs. (4)–(6) and (11) in (10) leads to a somewhat complicated algebraic expression for the integrated intensity of illumination of the sample powder as a function of the dimensions and positions of the collimating apertures and the dimensions of the focal spot. Because of the complexity of the fully substituted Eq. (10) it is more convenient to evaluate it for specific experimental cases by first evaluating y_3, y_4, y_5, and I_4 and then substituting the numerical values in Eq. (10).

It would be tedious and unnecessary to reproduce in full here the complete analyses for cases A, C, and D. Suffice it to say that by a procedure very similar to that described above for case B the appropriate integrated intensity expressions can be derived. The expressions for all four cases are summarized herewith:

Case A.

$$I=2I_m\frac{x_3}{x_2}y_2. \qquad (12)$$

Case B.

$$I=I_m(y_3+y_4)-I_4(y_3-y_5), \qquad (10)$$

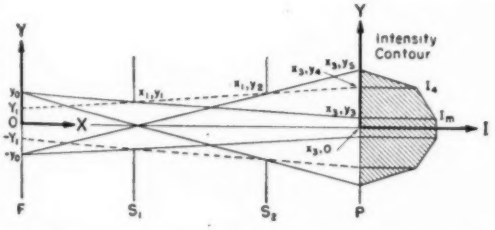


FIG. 4. The geometry of mathematical case B. The hatched contour depicts the intensity of sample irradiation as a function of the coordinate y.

TABLE I. Values of the x and y parameters for four typical experimental situations.

Parameter	Values of the parameters in mm (1)	(2)	(3)	(4)
x_1	60	60	60	60
x_2	100	100	100	100
x_3	150	150	150	150
y_0	0.500	4.000	0.500	4.000
y_1	0.375	0.375	2.750	2.750
y_2	0.375	0.375	2.750	2.750

TABLE II. Quantities computed in the evaluation of the integrated intensities for the four experimental situations of Table I.

Quantity	(1)	(2)	(3)	(4)
Y_1	0.375	0.375	2.750	2.750
Y_2	0.625	0.625	4.580	4.580
Y_3	1.500	1.500	11.000	11.000
Mathematical case	B	D	A	B
y_3	0.187	—	—	0.870
y_4	0.375	—	—	2.75
y_5	0.812	—	—	6.13
I_4	$0.875I_m$	—	—	$0.844I_m$
Calculated intensity in I_m units	1.11	0.26	8.25	8.06
Relative intensity (rectangular apertures)	13.4	3.2	100.0	97.5
Relative intensity (circular pinholes)	10.5	2.5	100.0	97.5

where

$$I_4=(1/2)I_m\left[\frac{x_2y_1-x_1y_2}{y_0(x_2-x_1)}+1\right]. \qquad (13)$$

Case C.

$$I=I_3(y_4-y_3)+I_4(y_3+y_5), \qquad (14)$$

where

$$I_3=I_m\frac{y_1}{y_0}\left(\frac{x_3}{x_3-x_1}\right) \qquad (15)$$

and I_4 is given by Eq. (13).

Case D.

$$I=2I_m\frac{x_3y_1y_2}{y_0(x_2-x_1)}. \qquad (16)$$

Equations (7) and (8), although not explicitly involved in the final intensity formulas, are essential to the derivation of Eqs. (12) and (16), respectively.

We may now return briefly to the observation made earlier that the Z dimensions of the apparatus could justifiably be disregarded in the present analysis. This action can be seen to be completely justified when the apertures S_1 and S_2 have the same width ($2z_1=2z_2$) and when at the same time the projected source dimension, $2z_0$, is equal to or less than $2z_1$, because this is simply mathematical case A applied to the Z dimension, and Eq. (12) shows I to be independent of y_0 (or z_0 in the present application). When z_0 begins to exceed z_1

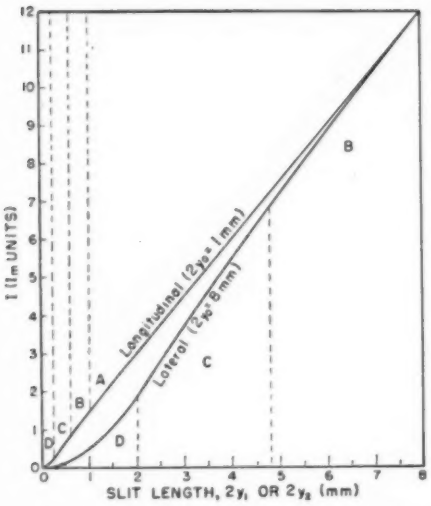


FIG. 5. Variation of integrated intensity, I, with slit length for a pair of identical rectangular slits of length $2y_1=2y_2$. The upper curve is for longitudinal and the lower for lateral view of the focus. Rectangular focal spot 1×8 mm; $\alpha=7°$; $x_1=60$, $x_2=100$, $x_3=150$ mm; slit width 0.75 mm. A, B, C, and D indicate applicable mathematical cases.

we are dealing with case B, but the application of Eq. (10) demonstrates that I diminishes very little until z_0 exceeds about $1.5z_1$.

For purposes of illustration we now apply these mathematical results to typical experimental conditions. Let $x_1=60$, $x_2=100$, and $x_3=150$ mm, and suppose the focal spot is a rectangle 1×8 mm in size. Furthermore, let us suppose that the focus is viewed at such an angle (7° in the present instance) that when observed in the longitudinal direction it appears as a square 1×1 mm in size. When viewed laterally, then, it will have the apparent dimensions 0.12×8 mm. Let us compare the integrated intensities of illumination for four important situations, including viewing the focal spot through: (1) two 0.75 mm pinholes longitudinally (customary arrangement using "end" windows), (2) two 0.75 mm pinholes laterally (using "side" windows), (3) two 0.75×5.5 mm slits longitudinally, (4) two 0.75×5.5 mm slits laterally. The proper values of the six x and y parameters for these four cases are given in Table I.

Table II lists the values of the various quantities which are calculated during the process of determining the integrated intensities for these four experimental situations. First, the quantities Y_1, Y_2, and Y_3 are computed, from which the appropriate mathematical treatment (case A, B, C, or D) is selected for each set of experimental conditions. Experimental cases (1) and (4) are found to fall in mathematical class B, and accordingly the quantities y_3, y_4, y_5, and I_4 are evaluated preliminary to the application of Eq. (10). Experimental cases (2) and (3) are found to fall in mathematical classes D and A respectively, for which no further quantities need be evaluated. Next the integrated intensity in I_m units is calculated using the appropriate equation, (10), (12), (14), or (16). The next row of data gives the relative intensities for the four cases referred to case (3) as 100. In the last row the relative intensities for the square apertures of edges $2y_1$ and $2y_2$ of cases (1) and (2) are converted to the equivalent values for circular apertures of diameters $2y_1$ and $2y_2$ by applying the factor $\pi r^2/(2r)^2=0.78$. The figures in the last row of Table II, then, give a direct comparison of the intensities to be expected when employing a pair of identical pinholes or a pair of identical rectangular slits.

The results given in Table II demonstrate clearly that a pair of pinholes is greatly inferior to a pair of slits of moderate length from the standpoint of irradiation intensity. Furthermore, the gain in intensity is found to be nearly the same whether the source be viewed longitudinally or laterally. These results agree with the experimental findings of Garrod[3] except for the precise extent of the intensity increases, which would hardly be expected to agree more closely because of differences in the respective experimental conditions. The second column of Table II shows a marked loss of intensity on viewing a rectangular focus laterally through a pair of pinholes with diameters much smaller than the length of the focus. This is, of course, the reason for the customary neglect of the "side" windows of the x-ray tube when only pinhole collimators are available.

The intensities provided by rectangular slits of various lengths are portrayed graphically in Fig. 5. The following typical conditions are assumed: $x_1=60$, $x_2=100$, $x_3=150$ mm; identical slits ($y_1=y_2$ and $z_1=z_2$); 1×8 mm rectangular focus viewed at 7°. The slit width ($2z_1$ or $2z_2$) is constant and equal to 0.75 mm. Figure 5 shows that the longitudinal direction of view is much preferable to the lateral for small values of $2y_1$ and $2y_2$, that is, when the slits are so short as to approach a square in shape. As the length of slit increases, the lateral intensity approaches the longitudinal, until when the slit length equals the length of the focus ($2y_1=2y_2=8$ mm in the present instance) the two techniques give equal intensities.

An analysis somewhat similar to the above has been carried out by Bolduan and Bear[5] for the special conditions inherent in low angle diffraction techniques. The results are applied to the design of collimators calculated to give optimum resolution and intensity.

[5] O. E. A. Bolduan and R. S. Bear, J. App. Phys. **20**, 983-992 (1949).

Average Electron Density Measurements by Low Angle X-Ray Scattering*

H. M. Barton, Jr. and R. Brill
Polytechnic Institute of Brooklyn, Brooklyn, New York, and Phillips Petroleum Company, Bartlesville, Oklahoma
(Received March 8, 1950)

According to Guinier the low angle x-ray scattering of particles should disappear when they are immersed in a fluid of equal electron density. This furnishes a way of measuring the electron density of powders, providing a suitable impregnating fluid can be found. Experiments of this type have been carried out with carbon black. A method of calculating the low angle scattering coefficient defined by Warren is given for impregnated samples. Measurements of the total low angle scattering intensity were made using a double crystal spectrometer and Geiger counter. No liquid was found that would give complete matching but approximate matching was obtained with phosphoric acid solution. The data extrapolate to show an electron density of $1.1N \pm 0.1N$ where N = Avogadro's number. This compares with a value of $0.99N$ as calculated from the helium immersion mass density measurement on the same sample.

I. INTRODUCTION

It has been stated by Guinier[1] that if particles of average electron density† ρ_e are suspended in a medium of electron density ρ_e', they will give low angle scattering as if they had electron density $(\rho_e - \rho_e')$ and were in a vacuum. This, of course, means that if there is no difference in electron density between the particles and their surroundings there will be no low angle scattering. This is analogous to the optical effect that particles immersed in a fluid of the same index do not scatter light.

Kratky, Secora, and Treer[2] investigated this experimentally for cellulose, finding in a qualitative way that the intensity of low angle scattering went to a low value when the particles were immersed in ethyl iodide, and that there was more intensity with fluids having greater or less electron density than the ethyl iodide.

Kratky and Wurster[3] extended the above experiments, measuring quantitatively the effect of different impregnating fluids on the small angle scattering of cellulose. They concluded that the intensity is proportional to the cube of the difference in electron densities. Their measurements were carried out using the photographic method. This procedure is rather complicated and requires many corrections, the magnitude of which may often be many times the measured value.

We used a more direct method, given by Warren.[4] By this method the total intensity scattered in the low angle region is measured by means of a double crystal spectrometer fixed at the maximum intensity angle. Warren defined a coefficient μ_{SA} which describes the reduction in primary beam intensity by a unit mass per unit area due to low angle scattering, in addition to that by the mass absorption; and showed it to be proportional to the integrated energy in low angle scattering. He derived the relation:

$$\mu_{SA} = \mu_m \frac{\ln(I_B/I_A)}{\ln(I_0/I_B)}, \tag{1}$$

where μ_m is the mass absorption coefficient for powders, I_A, I_B, and I_0 are the relative intensities measured with the sample in position A between the crystals, in position B in front of the Geiger counter, and out of the beam, respectively. The particles were in air.

II. SCATTERING COEFFICIENTS FOR IMPREGNATED SAMPLES

For samples impregnated with an ordinary liquid of mass absorption coefficient μ_i, we may set up the equation for the intensity at position A between the crystals

$$I_A = I_0 \exp[-\mu_i m_i - (\mu_p + \mu_{SA}) m_p], \tag{2}$$

where m_i is the mass per unit area of the impregnating liquid in the beam, μ_p is the mass absorption coefficient for powders of the particles, and m_p is the mass per unit area of the particles.

In position B in front of the Geiger counter where the low angle radiation is also registered:

$$I_B = I_0 \exp[-\mu_i m_i - \mu_p m_p], \tag{3}$$

and also we have for the total mass per unit area:

$$m_t = m_i + m_p. \tag{4}$$

Combining these three equations we get

$$\mu_{SA} = \frac{(\mu_p - \mu_i)\ln(I_B/I_A)}{\ln(I_0/I_B) - \mu_i m_t}, \tag{5}$$

which reduces to the Warren equation when μ_i is negligible. In the case where $\mu_p = \mu_i$, Eq. (5) becomes indeterminate, but this indeterminancy may be re-

* This is a condensation of a thesis presented by H. M. Barton in partial fulfillment of the requirements for a Masters degree at Polytechnic Institute of Brooklyn.

[1] A. Guinier, J. chim. phys. **40**, 133 (1943).

† $\rho_e = (n/M)\rho N$ where n is the average atomic number, M is the average atomic weight, ρ is the mass density, and N is Avogadro's number.

[2] Kratky, Secora, and Treer, Zeits. f. Elektrochemie **48**, 587 (1942).

[3] O. Kratky and A. Wurster, Zeits. f. Elektrochemie **50**, 249 (1944).

[4] B. E. Warren, J. App. Phys. **20**, 96 (1949).

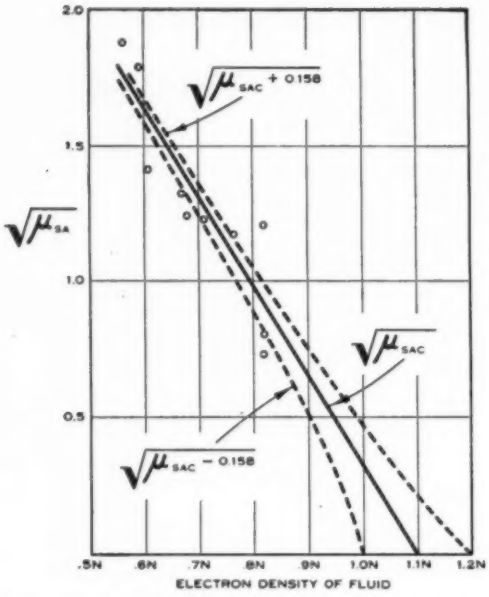


FIG. 1. Average electron density of impregnating fluid *vs.* square root of low angle scattering coefficient. μ_{SAC} is least squares calculation.

moved, giving

$$\operatorname{Lim}_{\mu_P \to \mu_i} \mu_{SA} = \frac{\mu_i \ln(I_B/I_A)}{\ln(I_0/I_B) - \mu_i m_i}. \tag{6}$$

This case will seldom be encountered, however.

III. RELATION OF SCATTERING TO RELATIVE ELECTRON DENSITY

The intensity of low angle scattering is generally considered to be proportional to the square of the number of electrons per particle, when the medium may be neglected. If the medium has appreciable electron density and if Guinier's statement holds, then the particles would scatter as if the electron density of the particles were diminished by that of the medium, and the effective number of electrons per particle would be $(\rho_e - \rho_e')V$ where V is the volume of a particle. The low angle scattering is thus seen to be proportional to $(\rho_e - \rho_e')^2$, or

$$\mu_{SA} = b(\rho_e - \rho_e')^2. \tag{7}$$

If this equation holds, then it will not be necessary to actually match the electron densities of the particles and fluid, but only to define the parabola (7) with sufficient accuracy.

IV. EXPERIMENTAL

To test this, a double crystal spectrometer was set up using calcites in the parallel position. The half-width at half-maximum of crystal B was about 19 seconds. Copper $K\alpha_1$ radiation was used, and the resultant radiation was measured with Geiger counter equipment removed from a Norelco x-ray spectrometer.

Experiments were made on impregnated carbon black. The fluids finally used were solutions of phosphoric acid. It was not possible to match the electron density of the carbon black with these solutions, so it was necessary to extrapolate to find the match point. It is difficult to find liquids with high electron densities which do not also have very high absorption. This requires reducing the thickness of the sample, but the sample must be kept thick enough to give sufficient low angle scattering. Some salts such as magnesium nitrate have suitable electron densities and mass absorptions. It proved possible to make pastes with molten salts but on solidifying the salts gave their own low angle scattering. Construction of a hot holder was not undertaken.

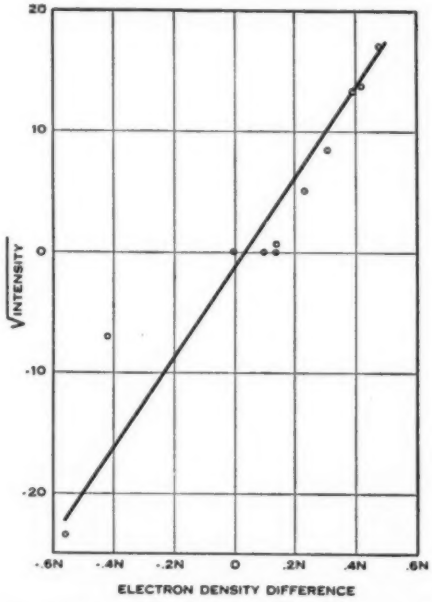


FIG. 2. Replot of data by Kratky and Wurster. Electron density difference between impregnating fluid and calculated electron density difference of hydrocellulose *vs.* square root of low angle scattering intensity. Intensity square roots are plotted as negative values for negative values of electron density difference.

V. RESULTS

Figure 1 shows the data plotted as the square root of μ_{SA} *vs.* the electron density of the fluid. The straight line is that obtained by least squares approximation to Eq. (7). It extrapolates to the axis at an electron density of $1.1N$. The average error expected for a measurement of μ_{SA} is:

$$p = \pm\left(\frac{\Sigma\Delta^2}{i(i-1)}\right)^{\frac{1}{2}}\left[1+\left(\frac{1}{2(i-1)}\right)^{\frac{1}{2}}\right], \tag{8}$$

where Δ is the difference between the least squares value and the experimentally measured value, and i is the number of measurements. For these data this comes out to be 0.158. The dashed curves were obtained by adding and subtracting this amount to the least squares curve and taking the square root. They are very close to

straight lines except near the intercept. The positive error curve does not actually go through the axis, but is shown extrapolated to get a rough idea of the error. These show an electron density for the carbon black of $1.1N \pm 0.1N$, corresponding to an electron density of $0.99N$ as calculated from the mass density determined by helium immersion of 1.98 g/cc.[5]

It would be highly desirable to obtain data closer to the point where the electron densities match. Experimentally this will require higher precision than was obtained in this work. It should be noted that the logarithm of the ratio of I_B to I_A occurs in Eq. (4). When there is little low angle scattering, this ratio is close to unity. Noting that when x is small the $\ln(1+x)$ is approximately x, it is easily seen that when the ratio is 1.01, a 1 percent error in measurement of the ratio will give a 100 percent error in the result. Measurements of intensities near the minimum of the parabola must therefore be made with great precision. Normally the extrapolation technique will give better results.

According to the paper by Kratky and Wurster,[3] the total intensity of small angle scattering should be proportional to $(\rho_e - \rho_e')^3$. Figure 2 shows their data replotted as the square root of the corrected intensities *vs.* $(\rho_e - \rho_e')$. We have the impression that a straight line fits with sufficient accuracy their results thus plotted to show the quadratic relation. Since it is not possible to obtain reasonable results with our data using the cubic assumption, we feel that the second power called for by theory is true.

We are indebted to Professor I. Fankuchen for several helpful discussions during the course of this work.

[5] L. R. Sperberg and H. M. Barton, Rubber Age **63**, No. 1, 1 (April, 1948).

A Removable Intermediate Lens for Extending the Magnification Range of an Electron Microscope

J. Hillier
RCA Laboratories, Princeton, New Jersey
(Received February 27, 1950)

A removable intermediate lens is described which extends the range of magnification of a conventional instrument to 25:1 without sacrificing the accessibility of the objective and projection lens polepieces. The distortion, sensitivity to changes in accelerating potential and reproducibility of calibration are studied and found to be equal to, or better than in the standard arrangement. The system provides two values of the magnification which are essentially independent of the accelerating potential.

I. INTRODUCTION

In conventional electron microscopes it has been customary to use an objective of fixed magnification in the neighborhood of 100× and single projection lenses of variable magnification with maximum values in the range 200 to 300. In other words the maximum electronic magnification by the entire instrument is usually in the range 20,000 to 30,000 diameters. Values greater than this are possible but have been used only occasionally owing to a number of technical difficulties including in particular, the high bombardment of the specimen necessary to provide sufficient image intensity. While the projection lenses are usually variable, more than a fourfold variation is rarely used owing to the introduction, at the lower values, of excessive image distortion and to an excessive limitation of the field of view. The minimum magnification attainable in the example cited above is in the range 5000 to 7500 diameters. However, these values are not sufficiently low for many uses of the electron microscope. For survey purposes or for correlation with the light microscope, magnifications in the range 500 to 1500 are desirable. It has been well known by those familiar with electron optics that the range of magnification and the field of view can be extended by the use of multiple projection lenses. Hillier[1] and Prebus[2] used double projection lenses each of which was weaker than a single conventional lens but together gave both a wide range of magnification (including values higher than those attainable with a single lens) and an increased field of view. This system did not find favor at the time because the lenses were not easily accessible and for some parts of the range excessive distortions were encountered. Marton suggested that lower magnifications might be included in the range by interposing a weak and variable lens between the objective and projector of a conventional instrument. More recently le Poole[3] has built an instrument in which he combined these two ideas introducing two lenses between the objective and projector. By properly activating one or two of these three projection lenses he is able to achieve any magnification he desires between zero and a maximum in the range 50,000 to 100,000 diameters. The adjustment can be made so that there is minimum distortion and satisfactory field of view at all magnifications.

In the design of electron microscopes easily accessible

[1] J. Hillier, U. S. Patent 2,323,328.
[2] A. F. Prebus, "Electron Microscopy," *Colloid Chemistry* (Reinhold Publishing Corporation, New York, 1944), edited by Jerome Alexander, Vol. V.
[3] J. le Poole, Philips Tech. Rev. **9**, 33, 1947.

TABLE I.

Tap no.	Total shunt resistance ohms	Instrument magnification at approximately 50 kv
1	400	7700
2	490	9000
3	580	10500
4	715	13000
5	880	14000
6	1060	15500
7	1330	17000
8	1720	18500
9	2720	20000
10	inf.	23400

polepieces have been considered of prime importance. This arises through the necessity of keeping the polepieces scrupulously clean. It has had the further advantage that a wide range of magnification could be obtained by the easy substitution of different lens combinations. For most work this has been found to be quite satisfactory. However, the convenience of a multiple lens system which enables the operator to survey his specimen at low magnification without breaking the vacuum, is not one to be overlooked.

In order to obtain the convenience of multiple lens operation without sacrificing polepiece accessibility in the electron microscope, a removable intermediate lens was designed which provides a range of magnification of 25:1 without changing polepieces. This new lens also reduces the distortion in the system, requires only a slight change in the electrical circuit of the RCA Model EMU, and provides the range of magnification in ten reproducible steps. The lens described was designed specifically for the standard RCA Model EMU in use in this laboratory.

II. DESCRIPTION OF REMOVABLE INTERMEDIATE LENS

Figures 1 and 2 are photographs of the complete unit. It is simply a lens spool consisting of two Armco iron end plates (one of which carries the central tube) threaded into a brass spacer which also acts as the lens-gap. The vacuum seal is achieved in the usual way by soft-soldering the threaded joints. The outer faces of the end plates are provided with projections which are machined respectively identical to the top and bottom ends of the intermediate chamber casting which it replaces in the instrument. The spool is surrounded by the usual sheath of Armco iron to complete the magnetic circuit. The spool is wound with insulated wire and must have a minimum resistance of 400 ohms. In our experimental model 9000 turns were wound on the coil which then had a resistance of 450 ohms. Connections to the coil were brought out into a shielded cable long enough to reach from the intermediate image position to the magnification control on the panel.

III. THEORY AND PERFORMANCE

The projection lens power supply in the Model EMU consists of a constant current regulator supplying 150 ma to the lens circuit and several regulator tube filaments. Variations in magnification are obtained by reducing the current in the projection lens; this being accomplished by a system of shunting resistors on a ten position step switch. (See Fig. 3) Table I gives the values

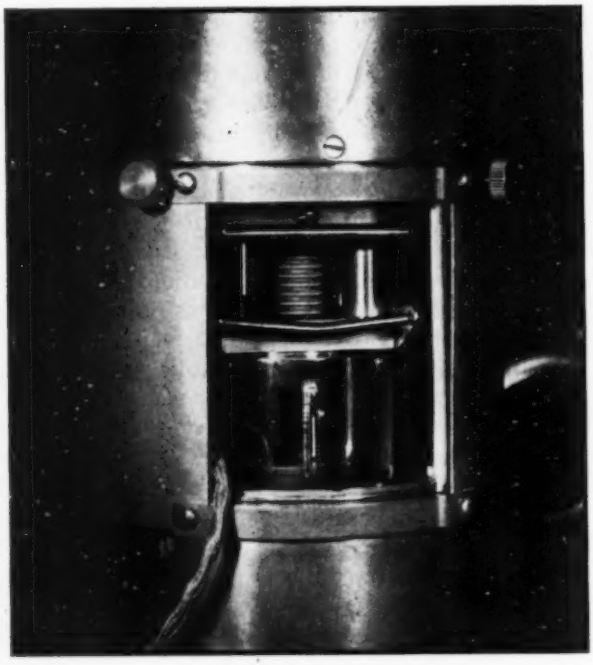


FIG. 1. A photograph showing the intermediate lens in position.

FIG. 2. A photograph showing the intermediate lens removed from the instrument.

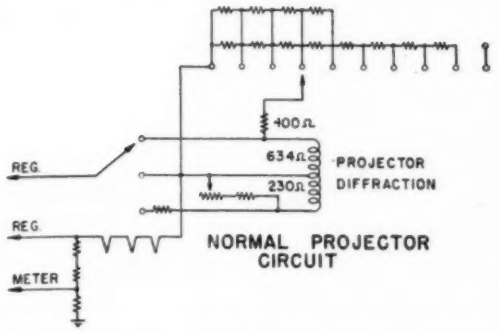


FIG. 3. A schematic diagram showing the circuit for the projection lens in the standard RCA Model EMU electron microscope.

of the shunt resistor and the approximate magnifications achieved in the instrument using the standard short spacer in the projection lens polepieces.

The modified circuit used with the removable lens is shown in Fig. 4. The 400-ohm resistor in the arm circuit of the shunt system is replaced by the coil winding and another shunt resistor of rather higher value. The actual value of this second shunt resistor is set experimentally to attain the lowest magnification possible consistent with a full field of view and minimum distortion. In our model this value was 1500 ohms giving a minimum magnification of 990.

The performance of this system can be derived from the following simple theory. Figure 5 is a simplified diagram of the optical system including the intermediate lens. The following notation and values are used: $u_1=-0.24$ cm, specimen to objective distance (fixed); v_1, image position for objective (cm); $u_2=v_1-d_1$, object position for intermediate lens (cm); $d_1=12.7$ cm, separation of objective and intermediate lens (fixed); $d_2=v_2=11.4$ cm, image position for intermediate lens (fixed); f_1, focal length of objective (cm); f_2, focal length of intermediate lens (cm); f_3, focal length of projection lens (cm); m_3, magnification of projection lens having values given in Table I; m, magnification of objective and intermediate lenses combined; $M=mm_3$, total instrumental magnification.

At a given projection lens setting the magnification of the objective and intermediate lenses combined will determine the change in the total magnification as compared to the standard arrangement since the circuit change has virtually not affected the current flowing through the projection lens. Thus, to determine the total magnification for each step, it is necessary only to calculate the magnification of the objective-intermediate lens system and multiply it by the known value of the magnification of the projection lens at the same step. Thus

$$m=v_1v_2/u_1u_2=v_1d_2/u(v_1-d_1) \tag{1}$$

but

$$1/f_2=1/v_2-1/u_2=1/d_2-1/(v_1-d_1) \tag{2}$$

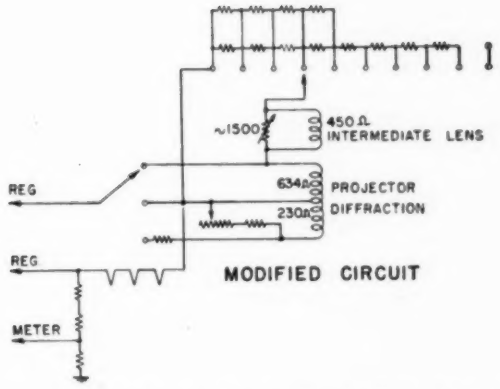


FIG. 4. A schematic diagram showing the circuit for the projection lens as modified for the insertion of the intermediate lens.

and

$$v_1=\frac{f_2(d_1+d_2)-d_1d_2}{f_2-d_2}. \tag{3}$$

Substituting in expression (1) gives

$$m=\frac{d_1+d_2}{u_1}\cdot\frac{d_1d_2}{f_2u_1} \tag{4}$$

which on substituting numerical values gives

$$m=100-(600/f_2)\text{ (very nearly).} \tag{5}$$

From the expression (5) it is apparent that the focal length of the intermediate lens is the controlling factor as far as the magnification m is concerned. Furthermore, as the power $(1/f_2)$ of the intermediate lens increases, the over-all magnification first decreases to zero then increases to the limit of the lens. Since the power of the lens increases with increasing current, in the circuit described, it will have maximum power for Step 1 on the magnification control. Since the lowest magnification is desired at this step and the current is known, the design of the coil can be established. The maximum intermediate lens power that will be required is such that $m=0$, that is, $f_2=6$ cm. Substituting numerical values in the equation for the focal length: $f_2=(KVD_2)/(n_2i_2)^2$ where $K=$instrumental constant (approximately 50); $V=50{,}000$ volts; $D_2=1.27$ cm; $n_2=$number of turns; $i_2=$current in amp., shows that the coil should have 715 amp. turns. Since the current to be expected is 0.093 amp. approximately 8000 turns are required. By using No. 30 wire, such a coil will have a resistance of approximately 400 ohms. The maximum power consumption is 3.45 watts.

From the above data it is now possible to calculate the total magnification to be expected from the system. In these calculations, the projection lens polepiece with the short spacer is used. Furthermore, the calculations apply to the model of the intermediate lens built and installed with a 1500 ohm shunt resistor. The results of these calculations are collected in Table II. The calculated

TABLE II.

Tap no.	Int. lens current (ma)	m	M calc. at 45 kv K adj. at step 1	M meas. at 45 kv.
1	75.3	11.8	990	990
2	68.7	26.3	2630	2340
3	63.4	37.1	4350	3890
4	56.9	49.5	7130	6350
5	50.6	60.1	9500	9300
6	45.2	68.3	11750	11650
7	38.8	76.5	14500	14600
8	32.3	83.8	17200	17500
9	22.5	92.0	20400	21600
10	0	100.0	26000	26900

TABLE III.

Exposure no.	Magnification
1	978
2	990
3	995
4	1015

and measured values of M are plotted on Fig. 6. It can be seen that the agreement is quite good in spite of the numerous assumptions and approximations inherent in the calculations.

Magnifications below 900 are not satisfactory because of excessive barrel distortion (see Fig. 7). Actually, at a magnification of 900 a single opening of the 200-mesh screen commonly used as a specimen support just covers the 2×2 in. image field. Thus, no increase in field is achieved until more than one opening is imaged. This requires magnifications below 500 which are beyond the range of the present system. They can be achieved however, by using the intermediate lens in combination with other projection lens polepiece arrangements. In every case the maximum magnification given by a given projection lens polepiece remains unchanged and the minimum is 1/25 (approximately) of the maximum. The image field in each case is that of the specific projection polepiece arrangement.

A point which should be mentioned in connection with the use of this lens is that the variation in focus with change in magnification is no longer negligible. This is particularly true at the lower values of magnification. The dependence of f_1 on f_2 is given by the expression:

$$\frac{1}{f_1}=\frac{f_2-d_2}{f_2(d_1+d_2)-d_1d_2}-\frac{1}{u_1}. \tag{6}$$

The substitution of numerical values shows that longer objective focal lengths are required for the lower magnifications. The increase for the entire range amounts to 7.1 percent most of which occurs in the lowest three steps. This corresponds to a reduction in objective current of approximately 3½ percent.

One other aspect of the optics of this system remains to be discussed, namely the distortion. Since for the range of values of f_2 considered the intermediate lens is always producing a real image of a virtual object, its spherical aberration results in a barrel distortion which becomes most apparent for those values of f_2 that cause the magnification m of the objective intermediate-lens system to approach zero. The projection lens, on the other hand, is always producing a real image of a real object and hence, as is well known, produces pincushion distortion. This is also more apparent at the lowest magnifications. Thus the distortion in the intermediate image is always opposite to that produced by the projector and tends to neutralize it.[4,5] It did not appear worth while to calculate the distortion since it was easier to measure it. Figure 7 shows the measured distortion at each magnification step. The distortion is expressed as a percentage increase in magnification, measured in the radial direction, at a point 2.5 cm from the center of the final image. It will be seen that for all practical purposes the system does not have excessive distortion. For comparison the distortion present in the final image produced by the standard projector lens with short spacer is plotted on the same figure.

IV. SOME PRACTICAL ASPECTS OF THE REMOVABLE INTERMEDIATE LENS

It is apparent from Fig. 6 that the relative rate of change of the total magnification M is very high for the lowest steps on the magnification control. For this reason, no attempt was made to design the intermediate lens for a specified minimum magnification. Instead, as already indicated, sufficient ampere-turns and resistance

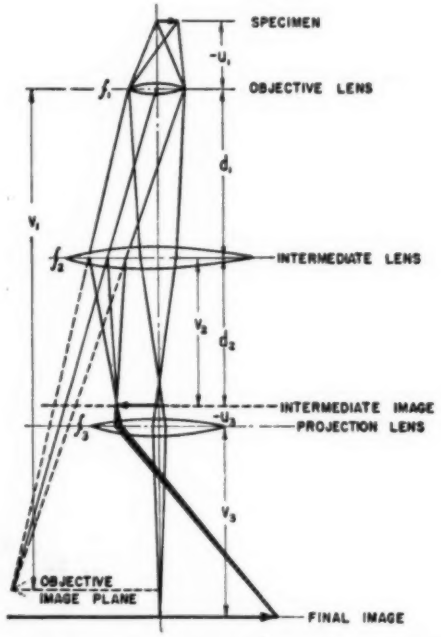


FIG. 5. A ray diagram illustrating the operation of the intermediate lens and the notation used in the text.

[4] J. Hillier, J. App. Phys. **17**, 411–419 (1946).
[5] J. Hillier, U. S. Patent 2,418,349.

were built into the coil to permit adjustment by external shunts. The theory shows, and has been verified, that the magnifications obtained for settings above Step 2 were relatively insensitive to the value of this external shunt. In our work the shunt was adjusted so that one opening of 200-mesh screen was seen in its entirety on the final screen at Step 1 on the control. At this adjustment the high barrel distortion of the intermediate lens made itself apparent by a "folding" into the center of the image of screen openings *outside* the field being examined. This difficulty was eliminated by placing an aperture (0.150–0.200 in. diameter) in the intermediate lens.

The high rate of change of the magnification at the lower settings makes the system very sensitive to hysteresis effects in either the intermediate or projection

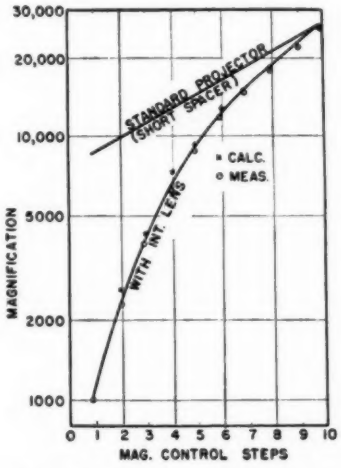


FIG. 6. A plot of the magnification against control position for the standard projector and for the standard projector in conjunction with the intermediate lens.

lenses. This was anticipated and particular care was taken in the choice of the iron used in the intermediate lens. When tested, however, it was found that the value of the lowest magnification was quite sensitive to the history of the lens. The extreme values encountered were 1110 and 990. The lower value was obtained by switching the projection lens circuit off and on several times. The higher value was obtained by leaving the circuit activated and turning the magnification control from Step 1 to Step 10 and back again several times. This manipulation showed that the effect was due to some retentivity in the iron of the projection lens since the two manipulations produce identical changes in the current flowing in the intermediate lens while they produce opposite changes in the current flowing in the projection lens.

In view of this situation, about which nothing could be done, some tests were made to check the reproducibility of the magnification at the lowest value. Four exposures were made of one field with the magnification at Step 1. Between the exposures the entire instrument

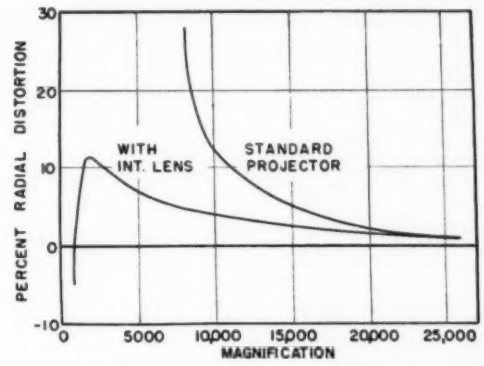


FIG. 7. A plot of percent radial distortion against magnification for the standard projector and for the standard projector in conjunction with the intermediate lens. The distortion is expressed as the percentage increase in magnification (measured in the radial direction) between the center and a point 2.5 cm from the center of the final image field.

was turned off and on and all the controls were changed in a random way. The image was then re-established and focused and the projection lens switch was turned off and on three or four times before making the exposure. The images obtained were then measured carefully. The results obtained are tabulated in Table III. It can be seen that the magnification had a mean value of 994 with a mean square deviation of 1.2 percent. Thus, while the system is sensitive to hysteresis effects, standardizing the magnetic history of the lens makes it possible to obtain very reproducible magnifications.

Since f_2 is proportional to the accelerating potential and the magnification given by Eq. (5) is a difference which approaches zero for the lowest step of the magnification control, it is apparent that the lowest magnifications are more sensitive to voltage changes than with the single projector. To check this, the percentage change in magnification which occurs when the accelerating potential was reduced from 50,000 volts to 45,000 volts was calculated for each step. The results are shown in Figs. 8 and 9; the former being a plot of the magnifications, the latter a plot of the percentage change in magnification. These calculations pointed out another important advantage of this system, namely that for Steps 4 and 5 (magnifications of 7140 and 9000) the *magnification of the system is essentially independent of*

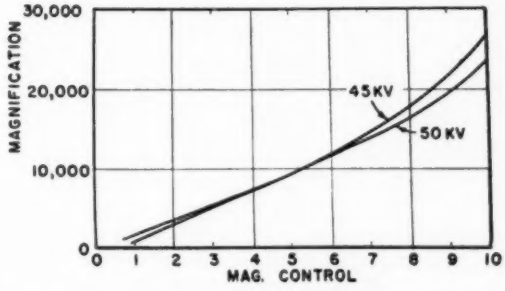


FIG. 8. A plot of the magnification calibration for the standard projector in conjunction with the intermediate lens for accelerating potentials of 45 kv and 50 kv.

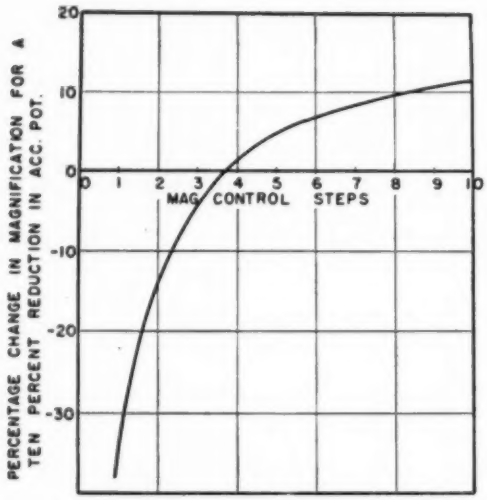


FIG. 9. A plot showing the percentage increase in magnification caused by a 10 percent reduction in accelerating potential.

the value of the accelerating potential. The potential change used in this example was 10 percent, which is many times greater than would be expected in normal use of the microscope, yet the change in magnification was only 1.5 percent for Step 5. Thus, if the specimen is placed accurately in the same position (within 0.001 in.) and the magnetic history of the intermediate and projection lens system is reproduced, the magnification calibration will remain accurate to within approximately 2 percent *for all time.* At all steps with the exception of Step 1, the dependence of the magnification on the accelerating potential (and hence on the state of the standard batteries) is the same or less than with a single projector.

The factors discussed in the preceding paragraph with regard to the dependence of magnification on the accelerating potential, also apply to changes in electron velocity caused by the absorption of energy in thick specimens. Thus, at Taps 4 and 5, chromatic change in magnification is minimized. This feature is particularly useful in the electron microscopy of sections of tissue. With uniformly thick 0.2-micron sections, the chromatic change in magnification has constituted a serious limitation of the field of view obtainable at the low magnifications used. To make the best use of the intermediate lens in the study of tissue sections, it is desirable to use a projection lens arrangement which gives a magnification of 700 to 1000 at Taps 4 and 5. The intermediate lens then becomes ideal for the examination of tissue sections.

A final point with regard to the practical use of the intermediate lens is the possibility of obtaining much lower magnifications for visual survey purposes by simply reducing the objective current. In experimental tests with thin specimens the defocusing could not be detected visually down to magnifications of 200 and was not serious to 100×. This does not apply to thick specimens such as tissue sections where the scattering greatly reduces the depth of field.

V. CONCLUSION AND SUMMARY

A removable intermediate lens has been described which extends the range of magnification in the direction of lower values without sacrificing the accessibility of the objective and projection lens pole pieces. Magnifications covering the range 900 to 24,000 are achieved without breaking the vacuum of the instrument. A complete field of view is obtained at all values and the distortion is in general less than with a conventional projector. The system provides one or two values of the magnification for which chromatic change in magnification is largely eliminated, thus providing the possibility of making an absolute and permanent magnification calibration.

While hysteresis effects in the existing projection lens cause the lowest values of the magnification to be sensitive to the magnetic history of the lens, it is shown that accurately reproducible magnifications can be obtained if the magnetic history is normalized.

The Wall-Quenching of Laminar Propane Flames as a Function of Pressure, Temperature, and Air-Fuel Ratio

RAYMOND FRIEDMAN AND W. C. JOHNSTON
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received February 13, 1950)

The quenching distance between plane parallel plates has been measured for propane-air flames at eight pressures ranging from 0.0832 to 2.77 atmos., and air-propane ratios ranging from 11 to 24. Other tests have been made at atmospheric pressure in which (a) the temperature of the quenching plates has been varied from 80° to 715°F, (b) the temperatures both of the inlet gas and the quenching plates have been varied from 80° to 545°F, and (c) a series of six different solid surfaces was tested. The minimum quenching distance is found to occur at an air-propane ratio between 13.5 and 14.0, and is proportional to the minus 0.91 power of pressure. Lean mixtures are found to be slightly less pressure-sensitive. When both the inlet gas and plates are heated, the minimum quenching distance is found to be proportional to the minus 0.5 power of absolute temperature. The quenching effect appears to be independent of the nature of the surface.

INTRODUCTION

THE quenching effect of relatively cold solid surfaces upon pre-mixed laminar gas flames has been explored by several workers, including Holm,[1] Lewis, von Elbe *et al.*,[2–4] Friedman,[5] and Forsyth and Garside.[6] The aim of this investigation is to extend the data for the propane-air system, with particular emphasis on the pressure effect. Quantitative measurements of the effect of surface temperature and of inlet gas temperature are also made for the first time, and a search for catalytic surface effects is described.

The importance of the quenching effect in Bunsen flame flash-back has been brought out by Lewis and von Elbe. The possibility exists that quenching plays a role in blow-off from a flame-holder, although this has not been conclusively demonstrated. Williams, Hottel, and Scurlock[7] found that heating of the flame-holder extends the stability limits considerably, which suggests that quenching by the cold anchor is of some importance. Particularly at low pressures, where the quenching distance is large, one might expect that incomplete combustion near cold burner walls is associated with quenching. Further, it appears reasonable that rich and lean inflammability limits and the low pressure limit of inflammability are governed by quenching effects. To illustrate this, the minimum pressure at which gasoline vapor-air may be ignited in a 2-in. tube is reported by Jones and Spolan[8] as 0.046 atmos. The extrapolated results of this investigation show the minimum quenching distance of propane-air at 0.046 atmos. to be 1.2 in. for plane-parallel walls, and this would be about 35 percent greater in a tube (or 1.6 in.), according to Fig. 6 of reference 3. This 1.6-in. quenching diameter for propane may be compared with the 2-in. tube in which gasoline ceases to propagate flame at this pressure, and it is seen that quenching must be important in burning near the low pressure limit. At atmospheric pressure also, the rich and lean limits as determined by upward propagation in tubes appear to depend upon quenching; data obtained by White[9] for five fuels (H_2, CO, CH_4, C_2H_4, C_2H_2) and air in 1-, 2-, and 3-in. tubes show a progressive broadening of both limits in each case with increasing tube size. The broadening of these limits may be thought of as extensions of the limbs of the familiar U-shaped curve of quenching distance against air-fuel ratio.

This investigation was undertaken, therefore, because of the general importance of the quenching phenomenon in combustion. Measurements were made by measuring the minimum spacing between plane-parallel plates through which flash-back can occur with a given fuel-air ratio, pressure, ambient temperature, and wall temperature. Propane was chosen as the fuel because of availability in pure form, ease in metering, and the

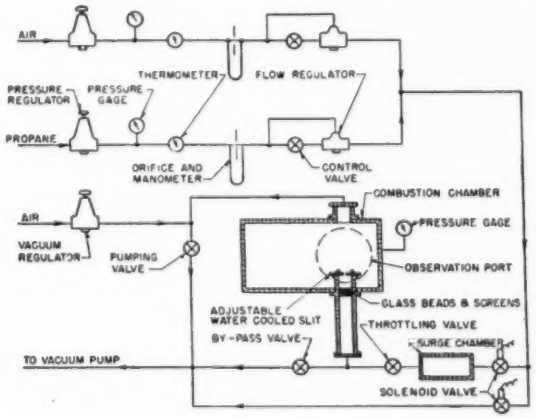


FIG. 1. Apparatus for measuring quenching distances at various pressures.

[1] J. M. Holm, Phil. Mag. **14**, 18 (1932); **15**, 329 (1933).
[2] Blanc, Guest, von Elbe, and Lewis, J. Chem. Phys. **15**, 798 (1947).
[3] Harris, Grumer, von Elbe, and Lewis, *Third Symposium on Combustion and Flame and Explosion Phenomena* (Williams and Wilkins, Baltimore, 1949), p. 80.
[4] Blanc, Guest, von Elbe, and Lewis, see reference 3, p. 363.
[5] R. Friedman, see reference 3, p. 110.
[6] J. S. Forsyth and J. E. Garside, see reference 3, p. 99.
[7] Williams, Hottel, and Scurlock, see reference 3, p. 21.
[8] G. W. Jones and I. Spolan, U. S. Bureau of Mines R. I. 3966 (1946).
[9] A. G. White, J. Chem. Soc. **125**, 2387 (1924).

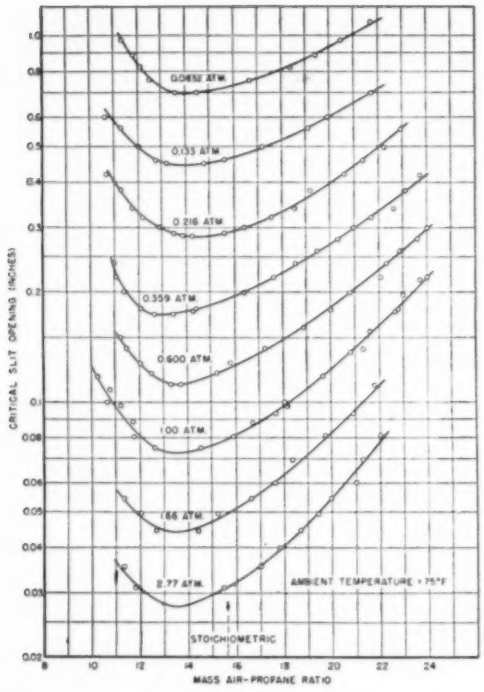


FIG. 2. The quenching of propane-air flames at various pressures.

similarity between propane and higher saturated hydrocarbons with regard to burning velocity,[10] minimum ignition energy,[3] flame temperature, and low temperature oxidation characteristics.[11]

EXPERIMENTAL TECHNIQUE

The procedure for determining the quenching distance consists in producing a flame stabilized on a rectangular port whose length-to-width ratio is always greater than 3.6. When the flow to the burner is quickly reduced to zero the flame will either be quenched by the plates or will flash back through the slit, depending on the air-fuel ratio of the mixture and the ambient conditions. The critical mixture is determined by testing a series of mixtures until a composition is found that separates a region of flash-back from a region of quenching. In general, there is a rich as well as a lean critical mixture corresponding to any particular set of conditions.

Variable-Pressure Flow Apparatus

The apparatus for measuring the quenching distance at various pressures is shown schematically in Fig. 1. A mixture of air and propane of controlled composition and pressure is supplied to the combustion chamber by a flow system. A thermocouple connected to a microammeter indicates the occurrence of flash-back, which can also be observed through a Pyrex window. The flow is reduced to zero by the action of the solenoid valves, surge chamber, and throttling valve. It was found that the sudden movement of the plunger in the solenoid valve was sufficient to disturb the flame, so that a surge chamber and throttling valve were necessary to smooth the flow during this period. The throttling valve was set to reduce the flow to zero in about 0.1 to 0.3 sec., although this setting was not found to be critical, and the data were reproducible over a fairly wide range of valve settings.

The vertical approach section to the burner is four inches in diameter and contains glass beads held in position by 60-mesh copper screens. The jaws of the slit are $\frac{3}{16}$-in. copper plates with water-cooling tubes attached. Slit widths are adjusted by thickness gauges.

The propane* used for these tests is reported to be 99.0 percent pure. The air is supplied at 80 p.s.i. g and contains 0.33 percent by weight water vapor. The effect of this concentration of water vapor upon the data appears to be negligible since the atmospheric curve of Fig. 2 agrees with previous work done with dry air.

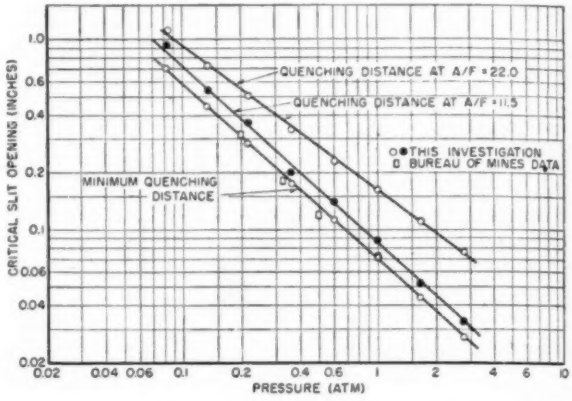

CRITICAL SLIT OPENING (INCHES)
PRESSURE (ATM)
QUENCHING DISTANCE AT A/F = 22.0
QUENCHING DISTANCE AT A/F = 11.5
MINIMUM QUENCHING DISTANCE
●THIS INVESTIGATION
□ BUREAU OF MINES DATA

FIG. 3. The dependence of propane-air flame quenching on pressure.

The flow meters were calibrated against a Sargent wet-test meter.

Variable-Temperature Flow Apparatus

The variable-temperature apparatus consists of identical air and propane flow systems. In each system the gas from a tank is reduced by a regulator and supplied to a critical-flow orifice.[12] The temperature and pressure at the orifice determine the density and hence the mass flow of gas. The pre-mixed gases are supplied to the burner through a hand-operated valve which is used for reducing the flow when testing. The burner consists of parallel copper plates with independent means for electrical heating of the plates and the gas. Tests showed that the results are independent of the plate thickness over the range from $\frac{1}{16}$ to $\frac{3}{16}$ in., as long as the inlet gases and the plates are at the same tempera-

[10] T. W. Reynolds and M. Gerstein, see reference 3, p. 190.
[11] K. Spence and D. T. A. Townend, see reference 3, p. 404.

* Matheson Company, East Rutherford, New Jersey.
[12] J. W. Andersen and R. Friedman, Rev. Sci. Inst. **20**, 61 (1949).

ture. Thermocouples are embedded in the plates for measuring the temperature of the quenching surface. Results indicate that the temperature gradient in the copper is negligible, being about 1°F per $\frac{1}{8}$ in. near the quenching surface when the flame is burning and the plates are water-cooled.

Factors Limiting Extent of Data

It is of interest to discuss the factors limiting the range of testing at the highest and lowest pressures and at the richest and leanest compositions. On the rich side, when air-propane ratios were below about 11, moving ridges appeared in the flame surface, often in a herringbone pattern. On further enrichment of the mixture, the flame would resolve itself into several individual cellular flames, each convex toward the unburned gas. Similar examples of polyhedral and cellular flames have been described in the literature.[13,14] When attempts were made to quench these flames, erratic results were obtained. Since these rich mixtures are characterized by very low burning velocities, it appears that sufficient time is available for surrounding gases to mix with the combustible gas beneath the flame surface.

On the lean side, it became progressively more difficult to ignite and stabilize the flames, the tendency for blow-off being very pronounced. At the lowest pressure the primary difficulty was ignition; it was found necessary to ignite each mixture at about 0.13 atmos. and then gradually reduce the pressure to 0.0832 atmos. for testing. This pressure represents substantially the lowest that can be obtained with the available pumping equipment. At the highest pressure, 2.77 atmos., and slit openings of 0.030 in. and less, testing was difficult because of excessive moisture condensation from the flame upon the plates.

In every case, data were taken over as wide a range as was feasible with the present apparatus.

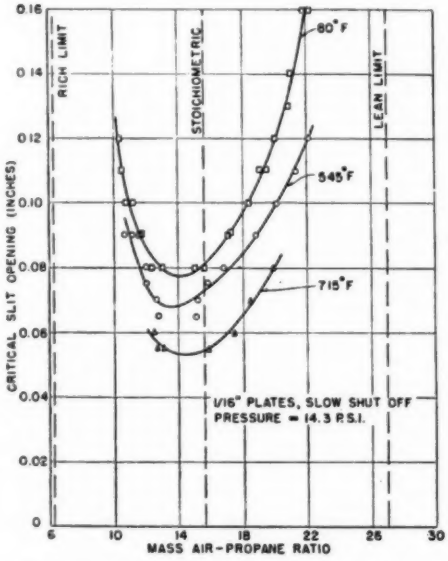


FIG. 4. Quenching of propane-air flames with various plate temperatures (gas initially 80°F).

[13] F. A. Smith and S. F. Pickering, Bur. Stand. J. Research **3**, 65 (1929).

[14] G. H. Markstein, J. Chem. Phys. **17**, 428 (1949).

TABLE I. The quenching of propane-air flames at various pressures, at 75°F.

Pressure (atmos.)	Slit opening (in.)	Critical mass air-propane ratio Lean	Rich	Pressure (atmos.)	Slit opening (in.)	Critical mass air-propane ratio Lean	Rich
0.0832	0.700	14.41	13.52	1.00	0.075	14.52	12.61
0.0832	0.760	16.64	12.48	1.00	0.0804	15.89	11.77
0.0832	0.822	18.32	12.06	1.00	0.088	16.70	11.71
0.0832	0.890	19.37	11.78	1.00	0.093	17.69	—
0.0832	0.980	20.40	11.29	1.00	0.098	18.15	11.20
0.0832	1.110	21.65	—	1.00	0.100	18.05	10.65
0.133	0.450	14.71	13.13	1.00	0.108	—	10.78
0.133	0.460	15.57	12.70	1.00	0.1175	19.63	10.25
0.133	0.500	17.13	11.98	1.00	0.137	20.80	—
0.133	0.560	19.05	11.26	1.00	0.140	21.30	—
0.133	0.600	19.87	10.58	1.00	0.157	21.60	—
0.133	0.700	21.70	—	1.00	0.177	22.70	—
0.216	0.285	14.22	13.82	1.00	0.180	22.80	—
0.216	0.290	15.54	13.47	1.00	0.197	23.00	—
0.216	0.300	16.38	12.89	1.00	0.2165	23.70	—
0.216	0.320	17.51	12.17	1.00	0.220	24.00	—
0.216	0.340	18.50	11.73	1.66	0.0455	14.41	12.68
0.216	0.380	19.15	11.25	1.66	0.0495	15.22	12.05
0.216	0.420	20.55	10.65	1.66	0.0545	16.63	11.32
0.216	0.460	21.35	—	1.66	0.060	17.63	—
0.216	0.500	22.20	—	1.66	0.0695	18.35	—
0.216	0.560	22.90	—	1.66	0.081	19.74	—
0.359	0.174	13.42	12.63	1.66	0.093	20.90	—
0.359	0.177	14.24	—	1.66	0.111	21.80	—
0.359	0.180	14.39	12.08	2.77	0.031	15.47	11.80
0.359	0.200	16.40	11.39	2.77	0.0355	17.01	11.32
0.359	0.220	17.60	11.04	2.77	0.040	17.81	—
0.359	0.240	18.51	10.96	2.77	0.0455	18.70	—
0.359	0.260	19.42	—	2.77	0.0495	19.42	—
0.359	0.280	20.30	—	2.77	0.0545	19.95	—
0.359	0.300	20.95	—	2.77	0.060	21.00	—
0.359	0.320	21.70	—	2.77	0.0695	21.30	—
0.359	0.340	22.60	—	2.77	0.081	22.04	—
0.359	0.380	23.10	—				
0.359	0.420	23.70	—				
0.600	0.112	13.70	13.30				
0.600	0.120	15.21	12.50				
0.600	0.128	15.79	12.02				
0.600	0.140	17.22	11.47				
0.600	0.150	—	11.23				
0.600	0.160	18.88	—				
0.600	0.180	20.00	—				
0.600	0.200	20.80	—				
0.600	0.220	22.05	—				
0.600	0.240	22.30	—				
0.600	0.260	22.85	—				
0.600	0.280	23.60	—				
0.600	0.300	24.00	—				

RESULTS

Precision and Accuracy

The data showing the effect of pressure upon quenching appear in Fig. 2. There are 118 points, each representing a group of bracketing tests; 78 percent of these points are within one percent of the curves, and 97 percent are within three percent of the curves. No points have been discarded. The points were not obtained in any regular sequence. Thus, the precision, or

TABLE II. The quenching of propane-air flames with various inlet temperatures, at 14.3 p.s.i. pressure (gas and plates at same temperature).

Temperature (°F)	Slit opening (in.)	Critical mass air-propane ratio Lean	Rich
80	0.080	14.95, 15.62	12.98, 12.32
80	0.090	17.27, 17.06	11.52, 11.71
80	0.100	18.35, 18.47	11.11, 10.68
80	0.110	19.15, 19.30	10.47
80	0.120	20.00	10.26
80	0.130	20.85	—
80	0.140	20.99	—
80	0.160	22.15, 21.75	—
260	0.070	14.84	13.14
260	0.075	16.08	12.20
260	0.085	18.08	11.38
260	0.090	18.85	11.10
260	0.100	19.81	—
260	0.120	21.57	—
410	0.060	14.84	13.35
410	0.065	16.08	12.53
410	0.075	18.37	11.46
410	0.085	20.13	10.80
410	0.090	20.64	10.52
410	0.100	21.75	—
410	0.110	22.83	—
545	0.057	15.23	13.45
545	0.058	16.60	12.94
545	0.065	18.56	11.54
545	0.075	20.13	11.01
545	0.080	21.24	10.66
545	0.085	21.71	10.33

reproducibility, of the test procedure is seen to be satisfactory.

An indication of accuracy was obtained by determining the quenching curve at 1 atmos. and 75°F both with the variable-pressure and the variable-temperature apparatus, these two flow systems having been independently calibrated. Also the burner geometry and method of flow reduction to produce flash-back were somewhat different in the two cases. Upon comparing the two curves at a given slit opening, the air-propane ratios never differed by more than two percent. The minimum slit openings differed by 5.5 percent.

The Bureau of Mines investigators[4] report data for minimum spacing of flanged electrodes for ignition in a static system with a spark of minimum energy. Their data for propane-air between 0.2 and 1 atmos. agree very well with the results of this investigation, as shown in Fig. 3.

Effect of Pressure and Composition

In Table I and Fig. 2, data are presented showing critical slit openings for quenching of propane-air flames from 0.0832 to 2.77 atmos., at 75°F ambient temperature. At each pressure, a U-shaped curve is obtained with the minimum at a mass air-propane ratio near 14 at the lower and 13.5 at the higher pressures. The stoichiometric air-propane ratio is 15.65. The minima are plotted against pressure in Fig. 3, a log-log plot, and a straight line is obtained, from the slope of which it follows that the minimum quenching distance is proportional to the minus 0.91 power of the pressure. It may be noticed in Fig. 2 that the pressure effect upon quenching is slightly smaller for the leaner mixtures. Upon plotting in Fig. 3 the quenching distance for an air-propane ratio of 22.0 against pressure, a straight line is again obtained, of reduced slope. The quenching distance is proportional to the minus 0.76 power in this case. For a mixture on the rich side of the minimum (air-propane ratio of 11.5), the minus 0.91 power is still valid.

The Bureau of Mines spark-quenching data for propane-air are seen to be consistent with these results both in magnitude and in pressure variation. Further, the Bureau of Mines spark-quenching data for methane-air between 0.1 and 1 atmos. reveal a minus 0.88 power pressure dependence when the minima are plotted in this fashion.

Effect of Temperature and Composition

In the first series of temperature tests at atmospheric pressure, the quenching plates were electrically heated to various temperatures while the inlet gas was maintained at 80°F. Results are shown in Fig. 4. It is seen that the quenching effect is diminished by increasing the plate temperature, as would be expected. In tests of this type, a certain unavoidable heating of the gas occurs as it flows through the slit. Rough measurements showed that the gas was heated to about 180°F when the plates were at 540°F. Because of this effect, the data of Fig. 4 are only semiquantitative.

Other tests were made at atmospheric pressure in which both the gas and the plates were heated to the same temperature, with results shown in Fig. 5 and Table II. The quenching effect is seen to become regularly less as the temperature is increased. For rich and stoichiometric flames, the quenching distance varies with about the

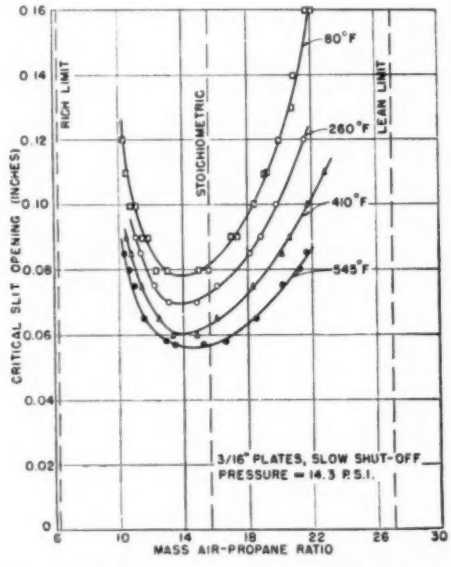


FIG. 5. Quenching of propane-air flames with various inlet temperatures (gas and plates at same temperature).

minus 0.5 power of absolute temperature, while for lean flames the negative exponent increases, being 0.85 at an air-propane ratio of 21.0.

Effect of Surface Coatings

A series of surface coatings on copper was investigated, with the gas and $\frac{3}{16}$-in. plates at 410°F. No measurable differences in quenching behavior were observed for either lean or rich mixtures, for any of the following types of surfaces: nickel, platinum, bright gold, roughened gold, potassium chloride, and water glass. This result indicates that neither the catalytic power of the surface nor its absorptivity for radiation significantly affects the quenching phenomenon.

Electromagnetic Field of the Conical Horn*

MARVIN G. SCHORR** AND FRED J. BECK, JR.
Yale University, New Haven, Connecticut
(Received January 20, 1950)

Maxwell's equations are solved for a perfectly conducting conical wave guide and the propagation coefficients of this guide are discussed. The field at the mouth of a finite conical horn is estimated and the radiation therefrom is calculated in integral form. These integrals are solved in series form for horns of small flare angle and moderate length and the results compared with experiment.

INTRODUCTION

THE use of electromagnetic horns for both matching purposes and enhanced radiation directivity in high frequency antennae was suggested by Southworth[1] and Barrow[2] in 1936. Since that time many observers have verified these properties and several attempts have been made to attain a satisfactory agreement between theory and experiment. The early workers calculated radiation phenomena from the scalar Kirchhoff-Huygens principle, but it was soon recognized that this formalism was inadequate, the major defect being in the assumption of continuity of the fields at the aperture from which the radiation ensues. Two rigorously correct methods are available for computing radiation from horn structures. They are the vector diffraction theory of Stratton and Chu,[3] and the assumed field technique recently restated by Schelkunoff.[4] A recent paper by Woonton, Hay, and Vogan[5] gives a good discussion of some of these matters.

The great majority of the existent theoretical work has been done for propagation in and radiation from uniform wave guide and sectoral horns. Buchholz[6] gave a rather complete analysis of electromagnetic fields in conical wave guide and also commented on a method of exciting such a guide. However he treated only the special case of azimuthally symmetric guided waves. Excellent experimental work is available in the literature on both conical[7] and pyramidal horns[8] and it is the purpose of this paper to show that the rigorous assumed-field technique, when applied to the conical horn, yields good agreement with experiment over a range of cone angles for which the radiation integrals may be satisfactorily evaluated.

FIELDS IN CONICAL WAVE GUIDE

Let us first consider the permissible electromagnetic fields inside a perfectly conducting conical wave guide of semi-infinite extent, filled with a dielectric of permeability μ and permitivity ϵ, ϵ being complex if the dielectric is lossy. The choice of a spherical (r, θ, ϕ) coordinate system is clearly indicated, and if the apex of the cone coincides with the origin of coordinates and the polar axis is the longitudinal axis of symmetry of the cone, then conical surfaces will coincide with surfaces of constant colatitude and the flare angle of the cone will be measured by the colatitude of its surface. The flare angle, so defined and henceforth used, is then one-half the total angle subtended by the cone. The cone is assumed excited near the apex in such a manner that the fields propagate positively radially outward; also the azimuthal reference axis is chosen so that a single trigonometric function describes the azimuthal variations of the fields. The permissible fields are well known and are given by:

Transverse-Electric Waves

$$
\begin{aligned}
E_r &= 0 \\
E_\theta &= Bj\omega\mu \cos m\phi h_n^{(2)}(kr) P_n^m(x) \\
E_\phi &= -Bj\omega\mu \sin\theta \sin m\phi h_n^{(2)}(kr) P_n^{m\prime}(x) \\
H_r &= B[n(n+1)/r] \sin m\phi h_n^{(2)}(kr) P_n^m(x) \\
H_\theta &= -B \sin\theta \sin m\phi [(1/r) h_n^{(2)}(kr) \\
&\qquad + k h_n^{(2)\prime}(kr)] P_n^{m\prime}(x) \\
H_\phi &= B[m/\sin\theta] \cos m\phi [(1/r) h_n^{(2)}(kr) \\
&\qquad + k h_n^{(2)\prime}(kr)] P_n^m(x);
\end{aligned}
\tag{1}
$$

* This paper is an abstract of a dissertation presented by the first named author (Marvin G. Schorr) for the degree of Doctor of Philosophy in Yale University, May, 1949.

** Now at Tracerlab, Inc., Boston, Massachusetts.

[1] G. C. Southworth, Bell Sys. Tech. J. **15**, 284 (1936).
[2] W. L. Barrow, Proc. I.R.E. **24**, 1298 (1936).
[3] J. A. Stratton and L. J. Chu, Phys. Rev. **56**, 92 (1939).
[4] S. A. Schelkunoff, Phys. Rev. **56**, 308 (1939).
[5] Woonton, Hay, and Vogan, J. App. Phys. **20**, 71 (1949).
[6] H. Buchholz, Ann. d. Physik **37**, 173 (1940).
[7] G. C. Southworth and A. P. King, Proc. I.R.E. **27**, 95 (1939).
[8] D. R. Rhodes, Proc. I.R.E. **36**, 1101 (1948).

Transverse-Magnetic Waves

$$
\begin{aligned}
E_r &= B[n(n+1)/r]\sin m\phi h_n^{(2)}(kr)P_n^m(x)\\
E_\theta &= -B\sin\theta\sin m\phi[(1/r)h_n^{(2)}(kr)\\
&\qquad +kh_n^{(2)\prime}(kr)]P_n^{m\prime}(x)\\
E_\phi &= B[m/\sin\theta]\cos m\phi[(1/r)h_n^{(2)}(kr)\\
&\qquad +kh_n^{(2)\prime}(kr)]P_n^m(x) \qquad (2)\\
H_r &= 0\\
H_\theta &= B(j\omega\epsilon m/\sin\theta)\cos m\phi h_n^{(2)}(kr)P_n^m(x)\\
H_\phi &= Bj\omega\epsilon\sin\theta\sin m\phi h_n^{(2)}(kr)P_n^{m\prime}(x),
\end{aligned}
$$

where the primed function always denotes the derivative with respect to the argument, and the wave number $k=2\pi/\lambda$, λ being the free space wave-length and $\omega^2=\mu\epsilon$. The $h_n^{(2)}(kr)$ and $P_n^m(x)$ are the spherical Hankel function of the second kind and associated Legendre function of the first kind respectively, and $x=\cos\theta$. The spherical Hankel function of the first kind does not appear since it represents a wave traveling in the negative radial direction, contrary to hypothesis; the associated Legendre function of the second kind also does not appear since it has a singularity on the polar axis which is contained in the region considered. The spherical Hankel function of the second kind has a singularity at the origin but this is of no consequence since, as will be seen later, the neighborhood of the origin is very highly attenuating so that no energy ever reaches it. The single-valuedness of the solutions in azimuth requires that m be an integer which we may take as positive. The colatitude boundary condition, namely $E_r=E_\phi=H_\theta=0$ at $\theta=\theta_0$, requires for TE waves,

$$[-\sin\theta P_n^{m\prime}(\cos\theta)]_{\theta=\theta_0}=0, \qquad (3)$$

and for TM waves,

$$P_n^m(\cos\theta_0)=P_n^m(x_0)=0. \qquad (4)$$

Hence n must be a root of the multibranched transcendental Eqs. (3) or (4) considered as functions of the order n of the associated Legendre function. In general n will be non-integral and a function of the flare angle θ_0 of the cone; we may consider only the positive roots. For small integral values of n, $0<n\leqq 10$, excellent tables of associated Legendre functions are available for determining the zeros of Eqs. (3) and (4)[9] and reasonably accurate interpolation to non-integral orders is possible. For large orders the zeros may be determined from the asymptotic expansions of these functions.

The solutions of Eqs. (1) and (2) then form a doubly infinite set corresponding to all positive integral m and all positive n giving zeros of Eqs. (3) and (4). Those modes actually present in the guide will of course be determined by the source of excitation.

It is convenient in discussing conical wave guide to adopt the cylindrical wave guide terminology since if we allow a given cross section of the cone to remain constant and let the flare angle approach zero, then in the limit the cone goes over into a hollow cylinder. The permissible modes in conical guide are hence designated TE_{mp} or TM_{mp} having m field variations in one azimuthal circuit and p field variations from the polar axis to the cone wall. The designation of propagation constants for conical guide does not, however, follow in the usual sense. In conical wave guide the fields are not periodic functions of the propagation coordinate r and propagation constant loses much of its intuitive meaning. However, in lieu of a better definition, a complex propagation coefficient is defined for conical wave guide, in the manner used by Barrow and Chu for sectoral horns,[10] by

$$\gamma(Q,r)=-1/Q(\partial Q/\partial r)=-(\partial/\partial r)(\log Q), \qquad (5)$$

where Q is any field component. In general γ will be a function of r and will be different for various field components. This definition, it may be observed, reduces directly to the usual definition of the propagation constant if Q is simple harmonic in r, and γ is then a true constant. For regions in which γ is not too rapid a function of radius useful information is still obtained.

In an entirely similar manner we may define a pseudo guide wave-length $\lambda_g(Q,r)=2\pi/Im\gamma$, and a pseudo phase velocity $v_p(Q,r)=\omega/Im\gamma$. From Eq. (1) for TE waves and the asymptotic expansions of the spherical Hankel function for large and small argument

$$
h_n^{(2)}(kr)=\begin{cases}
\left(\dfrac{\pi}{4}\right)^{\frac{1}{2}}\dfrac{1}{(n+\frac{1}{2})!}\left(\dfrac{kr}{2}\right)^n\\
\qquad +j\left(\dfrac{\pi}{4}\right)^{\frac{1}{2}}(n-\frac{1}{2})!\left(\dfrac{2}{kr}\right)^{n+1}, & kr\ll 1 \qquad (6)\\
\dfrac{1}{kr}\exp\left[-j\left(kr-\dfrac{n+1}{2}\pi\right)\right], & kr\gg 1
\end{cases}
$$

where $(n+\frac{1}{2})!$ has been written for the gamma function $\Gamma(n+\frac{3}{2})$, we may obtain the parameters of Table I which gives the propagation coefficient, the pseudo guide wave-length, and the pseudo phase velocity for the E_θ and H_ϕ components. Following Barrow and Chu we designate $kr\gg 1$ as a transmission region and that for which $kr\ll 1$ an attenuation region. In the transmission region the waves behave asymptotically as spherical waves in a free medium while in the attenuation region the behavior is similar to that in a uniform cross section guide near the cut-off wave-length.

These results can be expected to have physical validity only if the propagation coefficient γ is not too severe a function of r, or in other words, if the radial space variation is almost periodic. In the region near the apex where this condition is not true two anomalies arise. First, the phase of $\gamma(H_\phi)$ in the attenuation region is negative, which in uniform guide indicates a wave traveling in the negative radial direction, contrary to

[9] Math. Tables Project, Columbia University Press, New York (1945).

[10] W. L. Barrow and L. J. Chu, Proc. I.R.E. 27, 51 (1939).

hypothesis. Second, the group velocity v_g, if computed from the usual relation $v_g = d\omega/d(I_m\gamma)$ turns out to be of the same form as the phase velocity and the product of group and phase velocities is not c^2, as is required for a dispersionless medium. These anomalies can be ascribed to the severe aperiodicity of the Hankel functions in the region near the apex. It should be emphasized that the physical significance of the propagation coefficients here developed should not be exaggerated; however the implications of the results are of value in understanding the phenomena occurring.

Additional light on the properties of conical wave guide may be obtained from the behavior of the Poynting vector and power flow in the guide. It is not difficult to show that the angular components of the Poynting vector S are always pure imaginaries whereas the radial component is in general complex. In the attenuating region the radial component is largely imaginary along with the angular components. These latter decrease as $1/r^3$ at large distances and soon reach a negligible value compared with the radial component which decreases only as $1/r^2$.

From Eqs. (1) and (2), and the asymptotic expansions of the Hankel functions, Eq. (6), it is readily observed that in the transmission region all transverse field components decrease with reciprocal radius while the radial components decrease as reciprocal radius squared. As the wave progresses down the guide it therefore becomes more nearly a TEM wave matched to free medium, as may be seen from the wave impedance. For TE and TM waves, respectively, the wave impedance is given asymptotically by

$$kr \gg 1, \quad \begin{cases} Z_E = (\mu/\epsilon)^{\frac{1}{2}} \\ Z_H = (\mu/\epsilon)^{\frac{1}{2}} \end{cases}$$

$$kr \ll 1, \quad \begin{cases} Z_E = (\mu/\epsilon)^{\frac{1}{2}} \left[\dfrac{\pi(2n+1)}{n(n-\frac{1}{2})!(n+\frac{1}{2})!} \times \left(\dfrac{kr}{2}\right)^{2n+2} + j\dfrac{4}{nkr} \right] \\ Z_H = (\mu/\epsilon)^{\frac{1}{2}} \left[\dfrac{\pi(2n+1)}{(n-\frac{1}{2})!(n+\frac{1}{2})!} \times \left(\dfrac{kr}{2}\right)^{2n} - j\dfrac{n}{kr} \right]. \end{cases} \quad (8)$$

In regions near the apex the wave impedances are largely reactive, inductive for TE and capacitive for TM waves. As the wave moves into the transmission region the reactive part decreases, the resistive part increases and approaches asymptotically the characteristic impedance $(\mu/\epsilon)^{\frac{1}{2}}$ of the free medium.

RADIATION FROM A CONICAL HORN

Let us consider a cylindrical wave guide feeding a concentric conical wave guide. We will assume the cylindrical guide is excited in its lowest mode, TE_{11}, as is usually the case. At the junction of the uniform guide with the cone there is a discontinuity in the slope of the conducting boundary and higher order modes are required to satisfy the boundary conditions, the fundamental mode being partially reflected and partially transmitted. In the cylindrical guide the higher order modes will be rapidly attenuated but the behavior in the conical guide depends on whether the waves are introduced in the attenuating or transmitting regions of the cone. Certainly the cone will be transmitting for the fundamental mode on purely physical grounds since the diameter of the cone cross section is always greater than that of the uniform guide. The higher order modes, however, may be in an attenuating region and we have seen that such modes have an attenuation coefficient linearly proportional to the mode order, n, of the wave. The distance that a given order mode travels in an attenuating region may be calculated in a strictly qualitative manner, if it is assumed that in conical guide any order wave is in an attenuating region until the cross-sectional diameter of the cone approximates the diameter of a cylindrical guide just permitting propagation of that order wave in the uniform guide. If the diameter of the uniform guide operating in its lowest mode is D_0, and if the diameter corresponding to the nth mode is D_n, then the distance of travel of the nth mode wave in an attenuating region of the conical guide of flare angle θ_0 may be taken as

$$d = (D_n - D_0)/(2\sin\theta_0).$$

Small flare angles result in long journeys through attenuating regions. Also small flare angles result in smaller discontinuities at the cylinder-cone junction and, hence, require less higher mode energy to satisfy the boundary conditions. These factors both markedly discriminate against higher order modes.

In the conical guide, then, there will be present TE and/or TM waves composed principally of a fundamental and contaminated with higher order modes whose magnitudes depend on the flare angle and length of the guide. As the wave propagates in the positive radial direction, Eqs. (1), (2), and (6) show that, asymptotically, the radial components decrease only as reciprocal radius. Hence, some distance down the conical guide the wave will be almost transverse electromagnetic with only a small radial component. As this occurs, the wave impedance approaches that of a wave in free medium. Cutting off the conical guide at this distance should not materially alter the fields anywhere inside the cone since it is essentially matched into free space; very little energy will be reflected from the mouth. The small radial component still present is just sufficient to close the loops of the field lines of the TEM wave in free medium. The open ended cone thus becomes a source of radiation.

The calculation of the radiation characteristics of such a source involves two approximations at the outset:

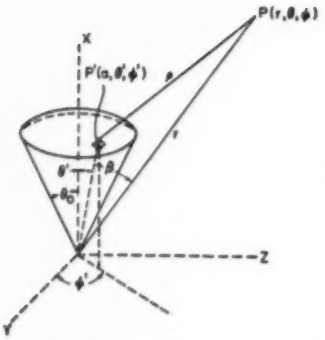


FIG. 1. Coordinate system for radiation calculations.

(1) the field at the mouth of the horn is assumed to be that which would exist at that cross section if the horn extended to infinity; and (2) there is assumed to be no field leakage around the rim of the mouth to the exterior surface of the horn. These matters have been discussed at length by others and will here be assumed valid provided the cone flare angle is not too great and its length is not too short.

The assumed field technique of Love,[11] MacDonald,[12] and later Schelkunoff[4] leads to the following expressions for the electric and magnetic intensities.

$$\begin{aligned} E&=k^2A+\nabla\nabla\cdot A-j\omega\mu\nabla\times F,\\ H&=k^2F+\nabla\nabla\cdot F+j\omega\epsilon\nabla\times A, \end{aligned} \tag{9}$$

where the vector potentials A and F are given by

$$\begin{aligned} F&=-\frac{1}{j\omega\mu}\int_{S'}\frac{[n\times E']ds'}{4\pi\rho},\\ A&=\frac{1}{j\omega\epsilon}\int_{S'}\frac{[n\times H']ds'}{4\pi\rho}. \end{aligned} \tag{10}$$

and E and H are the fields at any point in space distant ρ from a surface element ds' on which a field E' and H' exist, n being the unit vector normal to the surface and brackets indicating retardation in the Lorentz sense. The TE field at the mouth of the horn, S', is given by Eq. (1), and Fig. 1 shows the coordinate system used. Note that primed coordinates refer to the cone mouth while unprimed coordinates refer to the field point $P(r, \theta, \phi)$. Since the integrands are vector quantities the change in direction of the unit spherical vectors must be considered. This is most readily accomplished by expressing all vectors in rectangular coordinates. Noting that for the distant field

$$\rho\doteq r-a\cos\beta,$$

and

$$\cos\beta=\cos\theta\cos\theta'+\sin\theta\sin\theta'\cos(\phi-\phi');$$

and defining B_E and B_H by

$$B_E=B\frac{a^2}{4\pi r}e^{-jkr}h_n^{(2)}(ka)$$

$$B_H=B\frac{a^2}{j\omega\epsilon 4\pi r}e^{-jkr}\left[\frac{1}{a}-h_n^{(2)}(ka)+kh_n^{(2)\prime}(ka)\right],$$

so that $B_H/B_E=-Z_H$ at the spherical cap $r=a$; then the rectangular components of F and A are given by:

$$F_x=-B_E\int_0^{\theta_0}\int_0^{2\pi}e^{jka\cos\beta}\sin m\phi'\times\sin^3\theta' P_n^{m\prime}(x')d\theta'd\phi',$$

$$F_y=B_E\int_0^{\theta_0}\int_0^{2\pi}e^{jka\cos\beta}\{mP_n^m(x')\cos m\phi'\sin\phi'+P_n^{m\prime}(x')\sin m\phi'\cos\phi'\sin^2\theta'\cos\theta'\}d\theta'd\phi',$$

$$F_z=-B_E\int_0^{\theta_0}\int_0^{2\pi}e^{jka\cos\beta}\{mP_n^m(x')\cos m\phi'\cos\phi'-P_n^{m\prime}(x')\sin m\phi'\sin\phi'\sin^2\theta'\cos\theta'\}d\theta'd\phi',$$

$$A_x=B_H\int_0^{\theta_0}\int_0^{2\pi}e^{jka\cos\beta}mP_n^m(x')\times\cos m\phi'\sin\theta'd\theta'd\phi', \tag{11}$$

$$A_y=-B_H\int_0^{\theta_0}\int_0^{2\pi}e^{jka\cos\beta}\times\{mP_n^m(x')\cos m\phi'\cos\phi'\cos\theta'-P_n^{m\prime}(x')\sin m\phi'\sin\phi'\sin^2\theta'\}d\theta'd\phi',$$

$$A_z=-B_H\int_0^{\theta_0}\int_0^{2\pi}e^{jka\cos\beta}\times\{mP_n^m(x')\cos m\phi'\sin\phi'\cos\theta'+P_n^{m\prime}(x')\sin m\phi'\cos\phi'\sin^2\theta'\}d\theta'd\phi'.$$

The azimuthal integrations are over an orthogonal interval and present no difficulty, but the colatitude integrations are not over an orthogonal range and the evaluation of these integrals does not seem possible, analytically, in their present form. For small flare angles approximations may be made which permit the evaluation of the integrals in a sufficiently converging series.

The horn is assumed excited in the dominant TE mode used by Southworth and King in an experimental study of the radiation patterns of conical horns with which the calculated patterns are compared. We take

$$\begin{aligned} \cos\beta&\doteq\cos\theta+\sin\theta\sin\theta'\cos(\phi-\phi')\\ P_n^1(\cos\theta')&\doteq\sin(\alpha\theta')P_n^1(\cos\theta_0)\\ P_n^{1\prime}(\cos\theta')&\doteq-\frac{\alpha\cos(\alpha\theta')}{\sin\theta'}P_n^1(\cos\theta_0), \end{aligned} \tag{12}$$

where $\alpha=\pi/2\theta_0$. The Legendre function approximation may be checked against tabulated values for integral n and is very good except for a small range where θ' is nearly equal to θ_0. The approximation in $\cos\beta$ involves neglect of the variation of $\cos\theta'$ over the aperture of

[11] A. E. H. Love, Phil. Trans. **A197**, 1 (1901).
[12] H. M. MacDonald, Proc. London Math. Soc. **E10**, 91 (1911).

TABLE I. Propagation coefficients in conical wave guide.

		E_θ component	H_ϕ component
$kr \gg 1$	γ	$1/r+jk$	$1/r+jk$
	λ_g	λ	λ
	v_p	c	c
$kr \ll 1$	γ	$\frac{n+1}{r}+jk\frac{\pi}{2}\frac{2n+1}{(n-\frac{1}{2})!(n+\frac{1}{2})!}\left(\frac{kr}{2}\right)^{2n}$	$\frac{n+2}{r}-jk\frac{\pi}{2}\frac{(2n+1)(n+1)}{n(n-\frac{1}{2})!(n+\frac{1}{2})!}\left(\frac{kr}{2}\right)^{2n}$
	λ_g	$\lambda\cdot\frac{2}{\pi}\frac{(n-\frac{1}{2})!(n+\frac{1}{2})!}{2n+1}\left(\frac{2}{kr}\right)^{2n}$	$\lambda\cdot\frac{2}{\pi}\frac{n(n-\frac{1}{2})!(n+\frac{1}{2})!}{(2n+1)(n+1)}\left(\frac{2}{kr}\right)^{2n}$
	v_p	$c\cdot\frac{2}{\pi}\frac{(n-\frac{1}{2})!(n+\frac{1}{2})!}{2n+1}\left(\frac{2}{kr}\right)^{2n}$	$c\cdot\frac{2}{\pi}\frac{n(n-\frac{1}{2})!(n+\frac{1}{2})!}{(2n+1)(n+1)}\left(\frac{2}{kr}\right)^{2n}$

the horn. The phase factor may now be written as

$$e^{jka\cos\beta}=e^{jka\cos\theta}\cdot e^{j\frac{1}{2}ka\sin\theta\sin(\theta'+\phi-\phi')}\cdot e^{j\frac{1}{2}ka\sin\theta\sin(\theta'-\phi+\phi')},$$

and the last two factors expanded in a Fourier-Bessel series of the form

$$e^{jz\sin y}=\sum_{p=-\infty}^{\infty}J_p(\gamma)e^{jpy}.$$

Setting

$$\gamma=\tfrac{1}{2}ka\sin\theta,$$

$$e^{jka\cos\beta}=e^{jka\cos\theta}\sum_{p=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}J_p(\gamma)J_s(\gamma)\cdot e^{j[(p+s)\theta'+(p-s)\phi-(p-s)\phi']}. \tag{13}$$

The neglect of the variation of $\cos\theta'$ over the mouth is thus seen to be a first order approximation in that terms of the second order and higher are neglected. It should be noted that the extent of the approximation is a function of the length, ka, of the horn and, for long horns, the neglected portion of the phase factor may have a considerable effect on the integrals, even though θ' varies over only a small range. The approximation is best for short horns and for co-latitudes far removed from the polar axis.

In the expansion, Eq. (13), the azimuthal coordinate ϕ' appears only in the trigometric form $e^{-j(p-s)\phi'}$ and this expansion when inserted in Eq. (11) is multiplied by other trigonometric functions of ϕ' and integrated over the orthogonal interval $(0, 2\pi)$. This will cause the vanishing of all terms of the double sum, Eq. (13), except those for which a definite constant relation exists between the indices p and s, and the double sum is reduced to a single sum over the index p. The vector potentials then become, from Eq. (11) with $m=1$,

$$F_x=-j\pi/2B_EP_n^1(x_0)e^{j(ka\cos\theta+\phi)}\times\sum_{p=-\infty}^{\infty}J_p(\gamma)J_{p-1}(\gamma)f_x(p,\theta_0),$$

$$F_y=-j\pi/2B_EP_n^1(x_0)e^{j(ka\cos\theta+2\phi)}\times\sum_{p=-\infty}^{\infty}J_p(\gamma)J_{p-2}(\gamma)f_y(p,\theta_0),$$

$$F_z=-\pi B_EP_n^1(x_0)e^{jka\cos\theta}\sum_{p=-\infty}^{\infty}\{\langle J_p(\gamma)\rangle_M^2f_z(p,\theta_0)+1/2e^{j2\phi}J_p(\gamma)J_{p-2}(\gamma)f_y(p,\theta_0)\},$$

$$A_x=\pi B_HP_n^1(x_0)e^{j(ka\cos\theta+\phi)}\times\sum_{p=-\infty}^{\infty}J_p(\gamma)J_{p-1}(\gamma)a_x(p,\theta_0),$$

$$A_y=-\pi B_HP_n^1(x_0)e^{jka\cos\theta}\sum_{p=-\infty}^{\infty}\{\langle J_p(\gamma)\rangle_M^2a_y(p,\theta_0)+1/2e^{j2\phi}J_p(\gamma)J_{p-2}(\gamma)a_z(p,\theta_0)\},$$

$$A_z=j\pi/2B_HP_n^1(x_0)e^{j(ka\cos\theta+2\phi)}\times\sum_{p=-\infty}^{\infty}J_p(\gamma)J_{p-2}(\gamma)a_z(p,\theta_0); \tag{14}$$

where the "f" and "a" functions are defined by

$$f_x(p,\theta_0)=(\alpha/2)\{g_1(2p+1,\theta_0)-2g_1(2p-1,\theta_0)+g_1(2p-3,\theta_0)\},$$

$$f_y(p,\theta_0)=-j(\alpha/4)\{g_1(2p,\theta_0)+g_2(2p-2,\theta_0)-g_1(2p-4,\theta_0)\},$$

$$f_z(p,\theta_0)=j(\alpha/4)\{g_1(2p+2,\theta_0)-g_2(2p,\theta_0)-g_1(2p-2,\theta_0)\},$$

$$a_x(p,\theta_0)=(\alpha/8)\{g_2(2p,\theta_0)-g_2(2p-2,\theta_0)\},$$

$$a_y(p,\theta_0)=-j(\alpha/2)\times\{\tfrac{1}{4}[g_2(2p+1,\theta_0)+g_2(2p-3,\theta_0)]-[g_1(2p+1,\theta_0)-g_1(2p-3,\theta_0)]\},$$

$$a_z(p,\theta_0)=-j(\alpha/2)\times\{\tfrac{1}{4}[g_2(2p-1,\theta_0)+g_2(2p-3,\theta_0)]+[g_1(2p-1,\theta_0)-g_1(2p-3,\theta_0)]\}, \tag{15}$$

and the "g" functions are given by

$$g_1(K,\theta_0)=\frac{\alpha e^{iK\theta_0}-jK}{K^2-\alpha^2}$$

$$g_2(K,\theta_0)=-\frac{4}{\alpha}\frac{Ke^{iK\theta_0}-j\alpha}{K^2-\alpha^2}. \tag{16}$$

The "g" functions and hence "f" and "a" converge as reciprocal summation index provided $|K|$ is greater

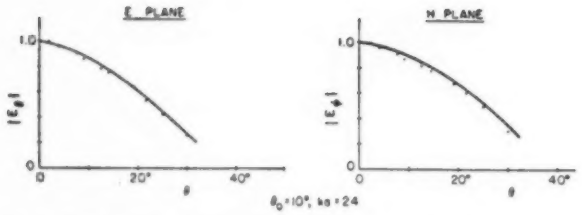


FIG. 2. Radiation pattern for conical horn of 10° flare angle. Solid curve is experimental and dotted points are calculated.

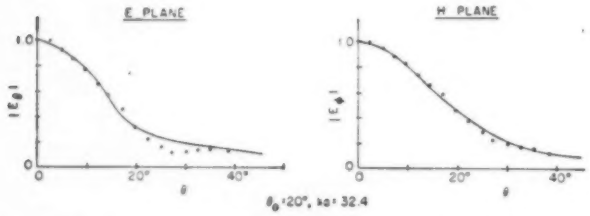


FIG. 3. Radiation pattern for conical horn of 20° flare angle. Solid curve is experimental and dotted points are calculated.

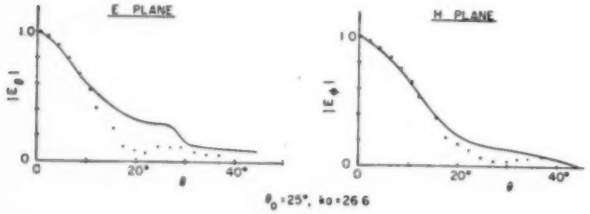


FIG. 4. Radiation pattern for conical horn of 25° flare angle. Solid curve is experimental and dotted points are calculated.

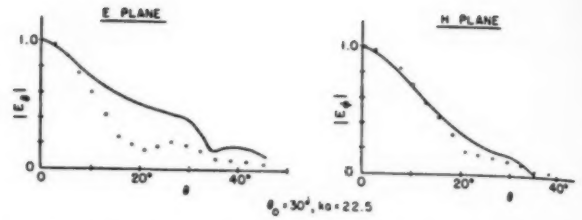


FIG. 5. Radiation pattern for conical horn of 30° flare angle. Solid curve is experimental and dotted points are calculated.

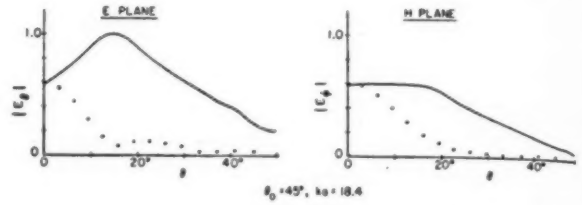


FIG. 6. Radiation pattern for conical horn of 45° flare angle. Solid curve is experimental and dotted points are calculated.

than α. The product of the two Bessel functions converges rapidly once the order p is greater than the argument γ. Hence the summations of Eq. (14) converge rapidly and, in general, only about γ terms must be included in the sum. These summations are over all integral values of p, positive and negative; however, the symmetry properties of the summands permits expression of the series as sums over only the positive integral indices. The only non-vanishing summations are those for which the total summand is an even function of the index of summation.

The vector potentials vary with radius as e^{-jkr}/r and thus if one ignores those radiation field components which vary faster than reciprocal radius, then Eq. (9) reduces to

$$E=k^2\frac{e^{-jkr}}{r}\{\theta_1[(\mu/\epsilon)^{\frac{1}{2}}F_\phi+A_\theta]-\phi_1[(\mu/\epsilon)^{\frac{1}{2}}F_\theta-A_\phi]\},$$

$$H=k^2\frac{e^{-jkr}}{r}\{\theta_1[F_\theta-(\epsilon/\mu)^{\frac{1}{2}}A_\phi]+\phi_1[F_\phi+(\epsilon/\mu)^{\frac{1}{2}}A_\theta]\},$$

where F and A are now the angular parts of the vector potentials, the radial parts being included in the e^{-jkr}/r factor. A receiving dipole oriented parallel to the E plane will then measure electric intensity patterns proportional to

$$E_{\theta E}\propto F_z-A_z\sin\theta+A_y\cos\theta$$
$$E_{\phi H}\propto -F_z\sin\theta+F_z\cos\theta+A_y$$

in the $E(\phi=0)$ and $H(\phi=\pi/2)$ planes, respectively.

The "g," "f," and "a" functions have been tabulated for various flare angles up to 45 degrees. If γ is restricted to values less than eight, then ten terms of the series are adequate. The field point colatitude at which numerical information is obtained is $\theta=\arcsin 2\gamma/ka$ and will be a function of the length of the horn. In general, it is found that longer horns give sharper radiation patterns and the increased resolution and decreased maximum field angle is not undesirable. However, no information on the small scale radiation at large polar angles is predicted. The greatest contribution to the radiation arises from the F_z and A_y components; F_z and A_z are negligible in all horns up to about 30 degrees flare angle.

Radiation patterns were calculated for horn flare angles from 10 to 45 degrees and horn lengths, ka, from 18.4 to 32.4. Horn parameters were chosen to agree with those used by Southworth and King[7] in an experimental study of conical horns fed by cylindrical guide at 15.3 cm free space wave-length. The flare angles used by them are total flare angles and hence double the flare angle θ_0 used here; the lengths of their horns were determined from the diameter of the aperture D and the excitation wave-length λ by

$$ka=(\pi D)/(\lambda\sin\theta_0).$$

The resulting patterns are shown in Figs. 2 to 6 in which the magnitude of the electric field is plotted as a function of the field point polar angle in the E and H planes. The solid curves are replots of Southworth and King's Fig. 10,[7] while the dotted points are normalized values calculated by the assumed field method. Ampli-

tude patterns are given here so as to compare directly with the experimental patterns. Only one-half the pattern is shown, since the theory is completely symmetrical, with respect to co-latitude, about the polar axis. There is little difference in the experimental curves between symmetrical halves but, where such differences exist, the average of the two halves was used to draw the solid curves.

The agreement between theory and experiment is seen to be good in the E plane for the 10° and 20° horns, poor at 25° and 30°, and very poor for the 45° horn. In the H plane the agreement is excellent for 10° and 20° horns, fair for 25° and 30°, and very bad for the 45° horn. The complete failure of the theory for the 45° horn is perhaps not surprising in the light of the small angle approximation made. Certainly it is not sound to neglect the variation of $\cos\theta'$ when the flare angle is as large as $\pi/4$. The possibility existed that the large discrepancy in the case of this 45° horn might be due to a considerable third harmonic excitation at the mouth because of the large discontinuities at the feed junction and the relative shortness of this horn. However, a qualitative calculation of the third harmonic contribution was unable to account for this behavior. The theory, in its present form, seems unable to yield a radiation pattern whose maximum amplitude is anywhere but on the polar axis, whereas experiment shows that it is quite possible to obtain maximum radiation in other directions. This effect has also been observed by Chu and Barrow[13] in the sectoral horn even after great care had been taken to eliminate all higher order modes. This peaking off the polar axis seems definitely to be a characteristic of large flare angle horns with which the present small angle theory is unable to cope.

Another anomaly in the theory is its prediction of varying horn length, at constant flare angle, on the pattern. The horn length, ka, enters only through γ, the argument of the Bessel functions, and hence the only effect of a change in ka is to alter the field point polar angle at which the calculation yields numerical results. An increase in length will sharpen the whole pattern, but no change should occur in the relative amplitudes of the side lobes and other structure. This prediction is not wholly confirmed by experiment in that, while the directivity tends to increase with length, lobes become more prevalent. This discrepancy again arises from the phase factor approximation where, it will be recalled, the neglect of the variation of $\cos\theta'$ can be justified only for small flare angles and moderate horn lengths. From the figures it is seen that the E plane directivity, i.e., that polar angle at which the amplitude falls to .707 of its maximum amplitude, is always somewhat greater than that in the H plane, the relation being approximately $E_{\theta E}=E_{\phi H}\cos\theta$. However, the H plane pattern is usually much smoother, with less side lobe, and falls to zero somewhat faster.

In conclusion it is perhaps worth while to point out that for pyramidal horns the same field Eqs. (1) and (2) apply provided only that the associated Legendre function of the first kind $P_n^m(x)$ is replaced wherever it occurs by

$$L_n^m(x)=aP_n^m(x)+bQ_n^m(x).$$

The radial part of these conical horn solutions apply directly to the pyramidal case and hence the various propagation factors computed for conical wave guide follow identically for pyramidal guide.

[13] L. J. Chu and W. L. Barrow, Trans. A.I.E.E. **58**, 333 (1939).

On the Diffraction of a Radar Wave by a Conducting Wedge*

R. B. Watson and C. W. Horton
Defense Research Laboratory, The University of Texas, Austin, Texas
(Received February 14, 1950)

Diffraction patterns of radar waves have been measured about the edge of a perfectly conducting wedge. Theoretical patterns have been calculated using an asymptotic solution suggested by Pauli. Good agreement is observed between experimental and calculated patterns. The thin wedge tested showed much similarity in diffracting properties to a suitable semi-infinite conducting screen. The results may have application to diffraction effects about wedge-shaped wings on high speed aircraft and missiles.

A SOLUTION to the problem of the diffraction of electromagnetic waves by a perfectly conducting wedge was obtained by Sommerfeld[1] and extended by MacDonald[2] and Carslaw.[3] More recently, an asymptotic solution was developed by Pauli[4] which allows for relatively easy calculation of this diffraction. It was pointed out in an earlier article[5] (*I*) on the diffraction of radar waves by a semi infinite conducting screen that optical measurements showed deviations from the theory due to the nature of the screen material and screen geometry. Wolfssohn[6] suggests that these optical results might agree better with theoretical calculations for a wedge, thereby accounting for the thickness of the diffracting material and the shape of the sharpened diffracting edge.

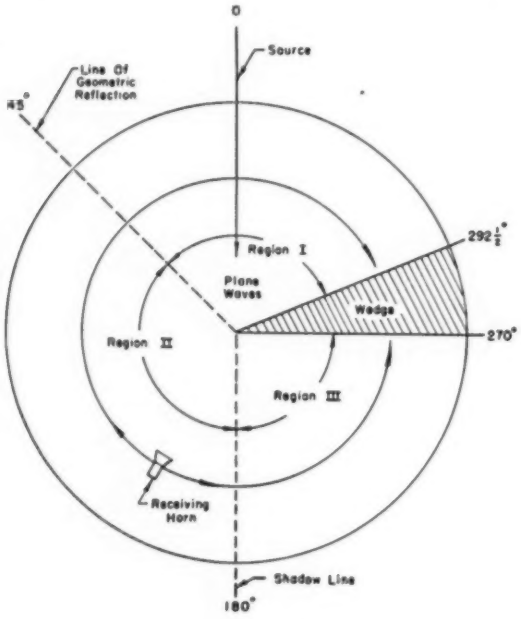


FIG. 1. Geometry of experimental arrangement.

The use of radar waves and metallic screens for obtaining diffraction measurements was justified fully in (*I*). The experimental arrangements described in (*I*) were used for the measurements on the wedge. The semi-infinite screen used in the earlier work was replaced by a hollow 22½° wedge made of two sheets of copper, each 30 in.×60 in.×$\frac{1}{16}$ in. thick. These sheets were supported in the general manner described in (*I*) except for the use of dural channel members placed between the support members for the two sheets in such a way as to form the rigid 22½° wedge structure with the two ends of the sheets brought together to form a straight edge.

Observations of the incidental diffraction about the upper and lower edges of the wedge, the free ends of the two sheets, and various supporting members were made by taking measurements with the receiving horn at 15 in. and 30 in. radii and at two positions along the diffracting edge: at the midpoint and at the upper edge of the screen. The patterns taken at 30 in. radius showed considerable incidental diffraction effects from other parts than the major diffracting edge, in agreement with measurements in (*I*); thus, data are presented for the 15 in. radius only.

Figure 1 shows the geometrical arrangement of the diffracting wedge with respect to the source and receiver. The source was a sufficient distance from the wedge that the incident waves were effectively plane. The receiving horn was rotated in a circle about the diffracting edge and was thus oriented to face the edge of the wedge at all times and to give maximum sensitivity for waves diffracted from this edge.

Figure 2 shows a polar plot of the received signal when the incident wave was polarized with the electric vector parallel to the diffracting edge. Figure 2 also includes a theoretical plot for the same condition. The wave field was resolved for calculation into plane and cylindrical waves, representing the incident and reflected waves, and the diffracted wave respectively, for the regions I, II, and III. (See Fig. 1.) For these regions, the diffracted wave was calculated by means of the approximate formula given in Eq. (13) of Pauli's paper.[4] In the regions near the shadow line and the line

* The work described in this paper was done at the Defense Research Laboratory, under the sponsorship of the Bureau of Ordnance, Navy Department, Contract NOrd-9195.

[1] A. Sommerfeld, Math. Ann. **45**, 263 (1894), and **47**, 317 (1896) (general method of solution).

[2] H. M. MacDonald, *Electric Waves* (Cambridge University Press, London, 1902), p. 186 (specific form for the wedge).

[3] H. S. Carslaw, Proc. London Math. Soc. **18**, 291 (1919).

[4] W. Pauli, Phys. Rev., **54**, 924 (1938).

[5] C. W. Horton and R. B. Watson, J. App. Phys. **21**, 16 (1950). This article will hereafter be denoted in the text by (*I*).

[6] *Handbuch der Physik*, Vol. 20, p. 278.

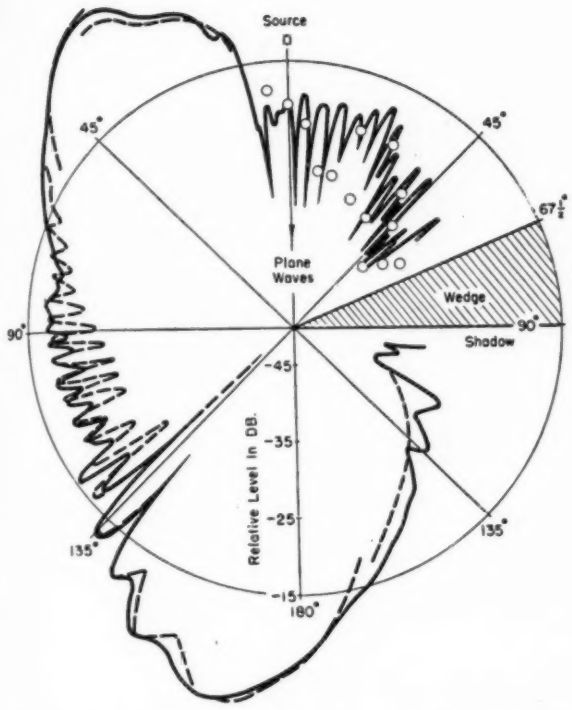


FIG. 2. Diffraction pattern for conducting wedge. *E*-vector parallel to diffracting edge. —— experiment; - - - and ○ theory.

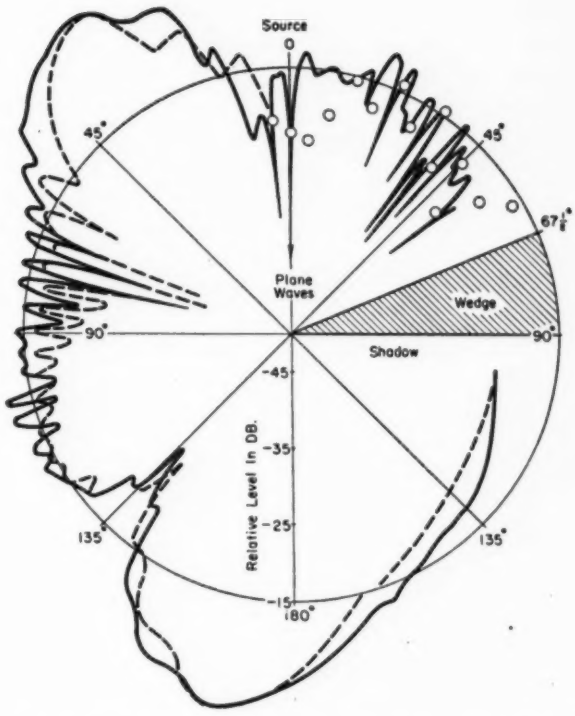


FIG. 3. Diffraction pattern for conducting wedge. *E*-vector perpendicular to diffracting edge. —— experiment; - - - and ○ theory.

of geometric reflection, the more accurate asymptotic series given as Eq. (35) in Pauli's paper was used.

Figure 3 gives a similar plot of the received signal when the incident wave was polarized with the electric vector perpendicular to the diffracting edge. Figure 3 further shows the computed pattern for this condition.

The agreement observed between theoretical and experimental patterns thus obtained again verifies the process of resolving the wave field into component fields. There appears a slight difference in the values for the two theoretical values for the angles removed 10° to 15° from the critical boundary lines. This difference arises because of the use of the two different types of approximate formulas in the various regions.

For comparison with optical results, the ratio of the amplitudes of the patterns for the electric vector parallel and perpendicular to the edge of the wedge has been plotted in Fig. 4 as a function of angle, for the shadow region only. The experimental curve agrees quite well with the theoretical curve in the region from 0° to 40°. Diffraction effects from other edges of the wedge and various surrounding objects become increasingly evident for larger angles. In line with Wolfssohn's suggestion that data taken for the wedge may fit optical data better than data taken for a semi-infinite screen, the optical data have been replotted here. There appears to be no significantly greater or less agreement than was observed in (*I*) for a semi-infinite screen.

The incident wave makes an angle of $-22\frac{1}{2}°$ with the normal to the forward sheet of the wedge. It is, therefore, of some interest to compare experimental measurements for the $22\frac{1}{2}°$ wedge with measurements made for a semi-infinite screen at an angle of incidence of $-22\frac{1}{2}°$. Such data are available in (*I*). An intercomparison shows only slight differences in detail. Similar agreement is obtained between experimental

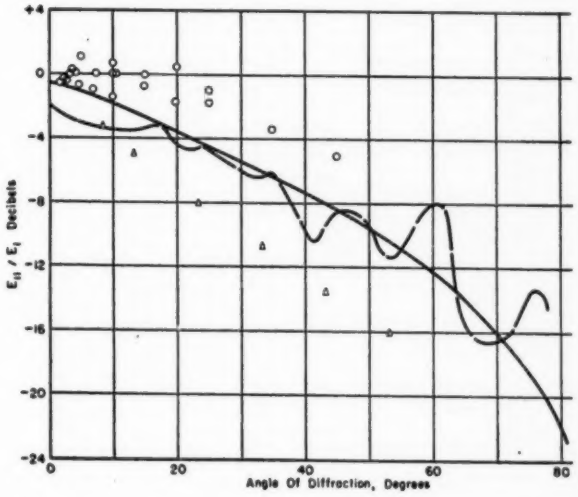


FIG. 4. Ratio of pattern amplitudes for electric vector parallel and perpendicular to diffracting edge. —— theory - - - experiment. Optical data: ○ Maey, Δ Jentzsch.

patterns for a quarter-infinite plane for an angle of incidence of $-22\frac{1}{2}°$ with the normal to the plane and patterns taken about the top edge of the $22\frac{1}{2}°$ wedge.

It is concluded that the diffraction caused by a thin wedge is not far different from the diffraction caused by the front surface of this wedge alone. Thus for such thin wedges, the rear sheet comprising the rear surface of the wedge is shown to be effectively shadowed by the front surface, and hence produces only small variations in the diffraction. These results may be of use in assessing diffraction about wedge-shaped wings on high speed aircraft and missiles.

Microscopic Observation of the Solidification of Small Metal Droplets

D. Turnbull and R. E. Cech
Research Laboratory, General Electric Company, Schenectady, New York
(Received March 6, 1950)

The solidification behavior of small metal (10 to 100 micron diameter) droplets has been observed on a high temperature microscope stage. An abrupt change in surface appearance and in the case of high melting metals a sudden brightening ("blick") accompanies solidification. The solidification temperatures observed for a collection of droplets may be widely distributed, but a significant, usually the major, fraction of the droplets supercool some maximum amount $(\Delta T_-)_{max}$ that is reproducible and characteristic of the metal. For many metals $(\Delta T_-)_{max} \approx 0.18$ times the absolute melting temperature. $(\Delta T_-)_{max}$ is not much changed by wide variations in the cooling rate and droplet size. The growth rate of metal crystals is very great so that the solidification rate of the droplets is controlled by the nucleation frequency.

INTRODUCTION

In another article[1] it was pointed out that if a metal specimen is broken up into a large number of isolated parts catalysts for crystal nucleation may be localized in a small fraction of the particles. Thus, in most of the particles crystal nucleation must take place homogeneously. In that investigation the solidification behavior of entire aggregates of particles was followed. An alternative procedure that is often advantageous is to remove a small number of particles from an aggregate and observe their solidification behavior microscopically. This is done by placing them in a stage recently designed by one of us[2] for microscopic observation at elevated temperatures. If only a small fraction of the particles contain nucleation catalysts there is a high probability that most of the ones selected for microscopic observation will be free of them. This principle has guided the present investigation.

A distinct advantage of the microscopic method of observation is its adaptability to the study of metals which melt at very high temperatures as well as medium and low melting point metals. Thus, it becomes possible by the use of one experimental technique to compare the solidification behavior of many metals.

For a long time[3] it has been known that small particles of liquid gold supercool significantly before solidification. A sudden brightening or "blick" occasioned by the liberation of the heat of fusion (recalescence) is known to accompany the solidification of the supercooled particle. This phenomenon is observed in the assaying of gold.

Mendenhall and Ingersoll[4] have observed this phenomenon in a number of other high melting metals including platinum, palladium, silicon, rhodium and iridium.

In their experiments metal particles 50 to 100 microns in diameter were supported on a Nernst glower in air or a CO_2 atmosphere, the solidification being observed microscopically. Particles of the same metal often solidified at widely different temperatures but the amount of supercooling, ΔT_-, was considerable and appeared to be roughly proportional to the melting point of the metal. However, the actual numbers for the maximum ΔT_- were given for only two of the metals, platinum and rhodium—370°C for both. These results of Mendenhall and Ingersoll were incidental to an investigation of another topic and their significance does not seem to have been realized.

In the current investigation the supercooling of many metals has been observed directly and the phenomenon studied in detail.

EXPERIMENTAL

In addition to its ease of adaptation to observation of metals having wide variety in melting temperature the microscopic method has the further advantage that particles isolated from one another on the hot stage do not require a film on their surface to prevent coalescence. It has been pointed out that[1] any crystalline protective film on the surface of a metal particle is undesirable in these experiments since it may catalyze crystal nucleation. On the other hand there are certain definite disadvantages of the method that must be recognized.

[1] D. Turnbull (to be published).
[2] R. E. Cech (to be published).
[3] A. D. Van Riemsdyk, Ann. Chim. Phys. **20**, 66 (1880).
[4] C. E. Mendenhall and L. R. Ingersoll, Phil. Mag. **15**, 205 (1908).

In producing aggregates of particles or in transferring them to the stage it is possible for nucleation catalysts to be picked up. Also, the particles, though free of crystalline films, must be supported on the stage and the possibility exists that the supporting surface or certain areas of it might catalyze crystal nucleation. In order to minimize this possibility, flakes of freshly blown quartz or pyrex glass were used as the supporting surface. These glass flakes, being amorphous, would not be likely to catalyze crystal nucleation. Another disadvantage of the microscopic method is that vibrations probably are more likely to promote solidification of small particles on a rigid supporting surface than particles in the interior of an aggregate. Sometimes these vibrations can be minimized by floating the particle in a suitable liquid flux.

MATERIALS

The source, purity, form, and method of preparation of particles are summarized in Table I. When the metal was supplied in the form of rod or wire small particles were made either by crushing with a mortar and pestle if brittle or by chipping with a sharp fragment of pyrex if ductile. Nickel particles and one sample of copper particles were formed from wire (fabricated from form listed in Table I) by means of a metallizing spray gun and the particles collected on a clean surface. These particles and those of other metals particularly susceptible to oxidation probably contained considerably more oxide than would be inferred from the purity listed. These metals were always given an *in situ* treatment with pure hydrogen in order to remove oxygen prior to observations of solidification. Different lots of some metals were obtained from two or three independent sources and the source designation used in the presentation of results is indicated in the table.

MICROSCOPE STAGE

Full details in the design of the microscope stage used in this investigation will be published in another paper.[2] A rough sketch is given in Fig. 1. The sample to be investigated was supported on a $\frac{1}{4}$-in. wide and 10 mil thick strip of molybdenum or platinum that was heated electrically.

Temperature was controlled by the amount of current passing through the ribbon heater and was varied by means of a variac transformer operated from a constant voltage source. By maintaining a constant power input into the heater it was possible to keep the temperature constant to $\pm 5°$ for long periods of time. The maximum temperature attainable was limited by the melting point of the heater metal. The stage is vacuum tight and its atmosphere (hydrogen, helium, or vacuum) is controlled by the indicated gas inlet. Temperature was calculated from the e.m.f. developed in an 0.005-in. diameter thermocouple welded to the underside of the heater. This e.m.f. was measured with a General Electric thermocouple potentiometer. The thermocouple was calibrated to read the temperature of the particles by accepting the known melting point of the metal as the standard. Deviation from the nominal temperature determined from standard calibration charts was usually +2 percent at the melting point. Solidification temperatures were corrected by assuming the deviation to be a linear function of temperature. A chromel-alumel thermocouple was used for temperatures below 1200°C and a platinum-platinum 10 percent rhodium thermocouple for temperatures greater than this. A new thermocouple was used for each specimen placed in the stage in order to prevent cumulative contamination from evaporation and condensation of the metals investigated. The error in the measured supercooling due to uncertainties in thermocouple and potentiometer calibrations and potentiometer readings is estimated to be ± 5 percent at most.

TABLE I. Source and purity of metals used in this investigation.

Metal	Source designation	Company	Purity (percent)	Form
Selenium		Johnson-Matthey	99.92	Shot
Bismuth	*A*	Eimer and Amend	99.999	Powder
	B	Johnson-Matthey	99.999	Rod
Lead	*A*	National Lead Co.	99.995	Ingot
	B	Eimer and Amend	99.95	Powder
Antimony		Johnson-Matthey	99.92	Ingot
Aluminum		Eimer and Amend	99.97	Powder
Germanium		Eagle-Picher Co.	99.99	Pellets
Silver		Handy and Harmon	99.97	Powder
Gold	*A*	Eimer and Amend	99.98	Powder
	B	Handy and Harmon	99.96	Wire
Copper		Amer. Smelting and Refining Co.	99.999	Rod
Manganese		Johnson-Matthey	99.99	Electrolytic sheet
Nickel		Int. Nickel Co.	99.98	Pellets
Cobalt		Johnson-Matthey	99.99	Powder
Iron		Nat'l Radiator Co.	99.99	Powder
Palladium		Amer. Platinum Works	99.95	Wire

PROCEDURE

Figure 2 shows the two types of experimental arrangements that proved most satisfactory. In (a) the heater has two bends or stirrups that were introduced to

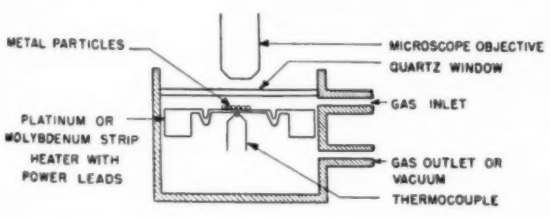


FIG. 1. Microscope stage with flat strip heater.

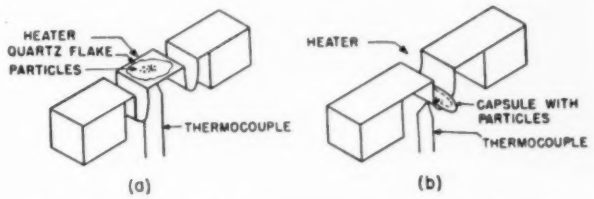


FIG. 2. Experimental arrangements used in investigation.

increase the effective length so that the temperature of the center horizontal section would be more uniform. This section supported a quartz flake on which a number of metal particles were placed for observation. In (b) a few metal particles were placed in the quartz capsule $\frac{1}{32}$-in. diameter by $\frac{1}{4}$-in. long which in turn was placed inside the single stirrup. A much larger number of particles could be observed at one time in arrangement (a) than in (b) in which the number was limited to 3–5. On the other hand, in (b) the temperature was very uniform along the length of the capsule and the calibration of the thermocouple more closely approximated that of the standard calibration charts. Also, (b) was much more satisfactory than (a) for observing volatile melts since in the latter arrangement the quartz window became fogged due to condensation of metal vapors on it. A new heater was used for each metal investigated in order to prevent contamination of the particles from condensation of metals accumulated in the heater during previous experiments.

An atmosphere of pure hydrogen, helium, or argon was maintained in the stage (excepting in the few instances that it was evacuated) during observations. Tank hydrogen was purified by passing it over a train consisting of a sulfuric acid bubbling tower and ascarite predryer, a copper chip furnace deoxidizer and a final dryer having dryerite and a liquid air trap. Tank helium or argon was forced through a purification train of two calcium chip furnaces in series operated at 550°C and 350°C.

In some of the experiments with metals melting at high temperatures Pyrex glass powder was mixed initially with the particles to provide a liquid coating during the solidification observations. It was thought that a liquid flux would cushion the particles against vibration and carry away or react with accidental crystal nucleation catalysts that might be present on the particle. However, it was demonstrated that the presence of a flux is not essential to the behavior to be described.

FIG. 3. Gold particles chipped from ingot. 25°C. 75×.

Melting and solidification of metals having melting points below 1000°C were inferred from the abrupt change of the surface condition of the particles that accompanied the processes. Upon initial melting, the particle lost its irregular shape and became spherical. Oblique illumination caused a bright spot or ribbon of light to be reflected from its mirror-like surface. When solidification occurred the rapid crystal growth in the particle caused a disruption of the smooth surface to one apparently having many sharp ridges and protrusions. These surface changes are illustrated in photomicrographs showing gold particles originally chipped from the ingot (Fig. 3), melted (Fig. 4), and solidified (Fig. 5). Melting and solidification temperatures were determined by association with the corresponding surface changes.

Solidification of metals having melting points greater than 1000°C was detected by the blick accompanying it. The initial melting of the particle permitted a thermocouple calibration as the particle became spherical. A further calibration was obtained by finding the maximum thermocouple e.m.f. that could be reached such that the particle did not blick upon cooling from the corresponding temperature. This e.m.f., which agreed within experimental error with that accompanying the initial shape change, was assumed to be that characteristic of the thermodynamic melting temperature. It was necessary to use filters for visual observations above 1100°C.

The melting and solidification temperatures of copper, gold, and germanium were measured by both the surface change and blick and the comparative results of the two methods are in complete agreement. Also, the maximum supercooling of bismuth particles (90°C) measured by observation of the surface changes is in close agreement with the value (90°C) determined dilatometrically by Turnbull.[1] In view of these facts there can be little doubt that the observed surface changes are caused by melting and solidification and that the blick observed in high temperature experiments is caused by solidification.

Particle diameters of the metals observed ranged from 10 to 100 microns excepting that some germanium particles were 400 microns in diameter. With the highest magnification that could be used, 240 diameters, it was not possible to detect the solidification temperature of particles smaller than 10 to 15 microns with certainty.

To determine the solidification temperature, the particles on the stage were cooled by decreasing the current through the strip heater. At intervals the cooling was interrupted and the thermocouple potentiometer exactly balanced and the e.m.f. recorded. The solidification temperature was taken to be that corresponding to the lowest exactly measured e.m.f. at which the particle was still liquid. Of the order of five minutes time usually elapsed in cooling from the melting point to the solidifi-

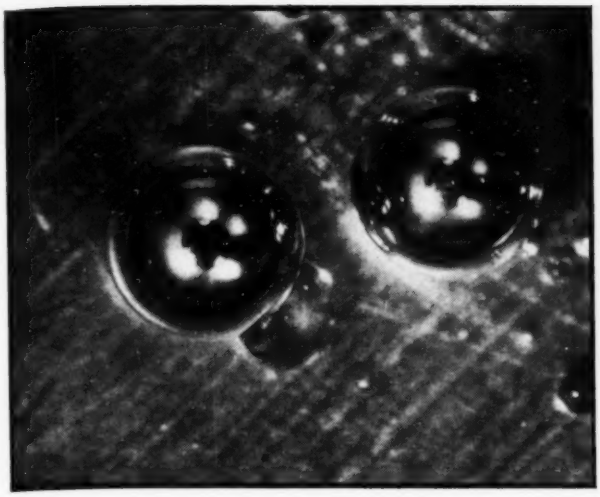


FIG. 4. Gold particles molten. 1070°C. 75×.

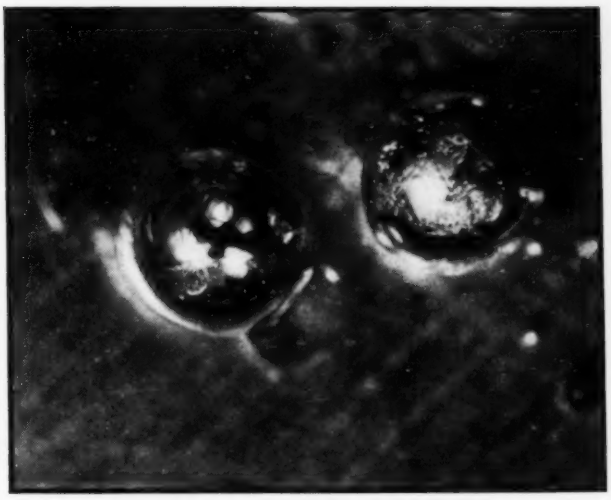


FIG. 5. Gold particles after solidification. 830°C. 75×.

cation temperature; most of this period was spent with the particles close to the solidification temperature. This cooling rate is rapid compared to that used in the dilatometric experiments but very small in relation to the maximum rate attainable (200°/sec.) with the apparatus. The temperatures of melting and solidification were checked several times for each particle or group of particles observed.

RESULTS

In general, the solidification temperatures of pure metal particles observed in a group ranged between the melting temperature, T_0, and some minimum temperature, T_{min}. The results could be represented in terms of a number distribution curve relative to solidification temperatures or degrees of supercooling ΔT_-, where $(\Delta T_-)_{max}=(T_0-T_{min})$.

Depending upon the metal and circumstances (e.g., surface films) the number distribution of solidification temperatures sometimes had a maximum at $\Delta T_-\rightarrow 0$ (this is not certain for metals studied by the blick procedure since a blick cannot be certainly detected for $\Delta T_-<25$), or at $\Delta T_-\rightarrow(\Delta T_-)_{max}$ or at some intermediate value of ΔT_-, and occasionally was fairly uniform so that no maximum was apparent.

It is an important fact that the solidification temperatures of a particular particle were not randomly distributed between $\Delta T_-=0$ and $(\Delta T_-)=(\Delta T_-)_{max}$ throughout successive melting-solidification cycles but remained fixed for at least several and usually all such cycles. Occasionally the solidification temperature, after remaining nearly constant for several cycles, did shift abruptly to some other value and remained there in a series of succeeding cycles. This behavior is analogous to that of the one g continuous gallium samples described in a preceding paper.[1]

As noted before the sizes of the particles observed generally were between 10 and 100 microns. From a statistical standpoint the solidification temperature of smaller particles tended to be less than that of larger ones. However, it is significant that the solidification behavior of particular particles often deviated consistently from this generalization. Thus, 10-micron particles were occasionally observed to solidify at the melting temperature while 50-micron particles solidified at $\Delta T_-\approx(\Delta T_-)_{max}$ throughout several melting-solidification cycles.

It appears reasonable to attribute the great differences in the solidification behavior of individual particles to the influence of minute quantities of accidental impurity particles. It is quite probable that colloidal particles suspended in the atmosphere and capable of serving as crystal nucleation catalysts become lodged in the quartz surface or on the surface of the metal particles during the assembly and operation of the microscope stage. (The experiments of Vonnegut[5] and Schaefer[6] have demonstrated that the concentration of colloidal material that can be effective in promoting the formation of snow is fantastically small.) Further, the abrupt change sometimes observed in the solidification behavior of particular particles can be attributed to the capture or dislodgement of a minute amount of impurity. *A priori* these impurities could either promote or inhibit solidification. However, there is much justification for the belief that such impurities promote crystal nucleation. This evidence has been summarized elsewhere[7] and will be discussed more fully in a following paper.[8] For the present purpose it will be assumed that minute impurities catalyze solidification. It follows that the only solidification temperature that could be characteristic of uncatalyzed solidification is $(\Delta T_-)_{max}$. For this reason only the maximum supercooling $(\Delta T_-)_{max}$

[5] B. Vonnegut, J. App. Phys. **18**, 593 (1947).
[6] V. J. Schaefer, Chem. Rev. **44**, 291 (1949).
[7] D. Turnbull, J. Chem. Phys. **18**, 198 (1950).
[8] D. Turnbull (to be published).

TABLE II. Summary of results of microscopic observations on supercooling of small droplets.

Metal	Source	Melting point °K (T_0)	Method of observation	Atmosphere	Flux	Particle size (microns)	Maximum supercooling $(\Delta T_-)_{max}$	$(\Delta T_-)_{max}/T_0$
Selenium		493	Surface change	Hydrogen	None	75	25[a]	0.051
Bismuth	*A*	544	Surface change	Hydrogen	None	10–15	90	0.166
	B	544	Surface change	Hydrogen	None	20–50	90	
Lead	*A*	600.7	Surface change	Hydrogen or vacuum	None	15–50	67	0.151
	B	600.7	Surface change	Hydrogen	NaOH+KOH	10–20	69	
Antimony		903	Surface change	Hydrogen	None	15–30	135	0.150
Aluminum		931.7	Surface change	Hydrogen	None	50–100	48	
					NaOH	50–100	130	0.140
Germanium		1231.7	Surface change+blick	Helium	None	15	235	0.184
					Pyrex	400	219	
Silver		1233.7	Surface change	Helium	Pyrex	20–40	227	0.184
Gold	*A*	1336	Surface change+blick	Helium or vacuum	Pyrex	20–50	221	0.172
	B	1336	Surface change+blick	Helium or vacuum	None	40–50	230	
Copper		1356	Surface change+blick	Hydrogen or helium	None	15–50	236	0.174
Manganese		1493	Blick	Hydrogen	Pyrex	50	308	0.206
Nickel		1725	Blick	Helium	Pyrex	50–100	319	0.185
Cobalt		1763	Blick	Helium	None	20–50	330	0.187
Iron		1803	Blick	Helium	None	30–100	295	0.164
Palladium		1828	Blick	Helium	None	30–100	332	0.182

[a] Corresponds to maximum supercooling that could be reached such that crystallization did not occur on re-heating.

that could be reproduced under various conditions will be listed in the results.

The results are summarized in Table II. Values of $(\Delta T_-)_{max}$ determined in different atmospheres or in vacuum agreed within experimental error. The $(\Delta T_-)_{max}$ listed is the average value. Absolute melting temperatures of the metals listed in column 3 are from Kelley's[9] compilation.

FRACTION OF PARTICLES SUPERCOOLING MAXIMUM AMOUNT

In general the major fraction of metal particles melting at temperatures more than 950°C (Ge through Pd) did not solidify until supercooled to or nearly to $(\Delta T_-)_{max}$. Particles of source (*A*) gold constituted an interesting exception to this behavior. Without any Pyrex flux none of the particles were observed to supercool more than about 40°C. In the presence of the flux the supercooling was larger and a significant fraction supercooled to $(\Delta T_-)_{max}$. On the other hand the major fraction of gold particles chipped off gold wire (source *B*) supercooled to $(\Delta T_-)_{max}$ either with or without flux. The difference in the behavior of these two samples appears to confirm the principle set forth in an earlier publication[1] that crystal nucleation catalysts are more likely to be absent in metal particles freshly formed by breaking up a massive piece of metal than in particles formed "separately" or that have been permitted to accumulate catalytic dust on their surfaces for long periods. Another factor that might account for the difference is that the particles of source *A* gold actually observed were coalesced from several of the original particles which by themselves were too small for their solidification to be detected by the microscopic method. Thus, in effect more surface having the possibility of holding catalytic dust went into source *A* particles than into source *B* particles.

The distribution of particles of low melting metals with respect to solidification temperature had some interesting aspects and will be considered separately.

EFFECT OF ATMOSPHERE

Whenever a high melting metal having an oxide stable at the melting point was studied pure hydrogen was passed over the heated particles for some time and the melting and solidification temperatures were observed concurrently. After this the stage was flushed with pure helium and solidification observations were made in this atmosphere. When results were obtained for both atmospheres those for helium have been put in the table for the reasons that (1) hydrogen is significantly soluble in some of the metals and (2) e.m.f.'s of the platinum platinum-rhodium couple are not reproducible when the junction is in a hydrogen atmosphere. However, the maximum supercooling in the two atmospheres did not differ by as much as the uncertainty of the measurements.

It is significant that the supercooling of copper par-

[9] K. K. Kelley, Bureau of Mines, Bulletin No. 393 (1936).

ticles fabricated with a metallizing spray gun increased by about 100° after having been treated in hydrogen at 1000–1100°C for about 15 min. This may indicate that Cu_2O that formed in the fabrication process catalyzes the formation of copper crystals. Further evidence in favor of this view is that copper particles chipped from the original rod that supercooled the maximum amount in hydrogen or helium did not supercool significantly when the atmosphere was changed to air at low pressure.

The solidification of iron particles that had been hydrogen treated, evacuated, and sealed off in a quartz capsule also took place at the same temperature as in a helium atmosphere.

ALUMINUM

As received, the aluminum particles had a rod-like form that persisted in the liquid state due to the rigidity of the oxide skin. When placed on a quartz leaf with no flux practically all of these particles supercooled 48 ± 2°C. The solidification range was so narrow that there was some temptation to assert that aluminum crystals nucleate spontaneously at $\Delta T_-=48$. However, in order to test the possibility that solidification was catalyzed by the oxide film, observations were made on solidification in a molten sodium hydroxide "flux." Under these conditions the particles spheroidized but a surface film (possibly sodium aluminate) remained. However, after heating 200° in excess of T_0, the particles supercooled 120° to 130°. This result appears to demonstrate that aluminum oxide does in fact catalyze the nucleation of aluminum crystals possibly by the mechanism suggested by C. S. Smith.[10] Of course there is no assurance that the film remaining in molten sodium hydroxide does not also catalyze solidification.

LEAD

Particles of lead placed on a Pyrex leaf in a hydrogen atmosphere solidified at temperatures ranging from the melting point to 60–65° below it at which only a minor fraction crystallized. As already noted, however, the behavior of particular particles was constant throughout successive melting-solidification cycles. When sodium hydroxide was used as a flux the fraction of particles solidifying at lower temperatures was greatly increased and a maximum supercooling of 69° was observed. On the other hand, in a flux consisting of a nitrate-nitrite mixture (56 percent KNO_3 and 44 percent $NaNO_2$) practically all the particles solidified at only 20–30° supercooling. A yellow film presumed to be lead oxide was present on the surface of the particles in the latter experiments. Thus, it seems likely that either lead oxide (PbO) or lead suboxide (Pb_2O) is a very effective catalyst for the crystallization of lead. Also it is interesting that the major fraction of source *A* particles supercooled to or nearly to $(\Delta T_-)_{max}$ with no flux when the stage was evacuated to a pressure of 0.01 mm of mercury.

The maximum supercooling obtained in these experiments is about the same order as the supercooling of the major portion of lead in dilatometric and thermal experiments described in a preceding paper but is more than 10° less than the maximum measured in dilatometric experiments. Also, excepting when a sodium hydroxide flux was used, the fraction of source *B* particles that supercooled to 55–65° was much smaller in the microscopic experiments than in the thermal and dilatometric ones. These results perhaps may be accounted for by the fact that the particles observed in the microscopic experiments were considerably larger on the average than those used in the dilatometric and thermal experiments and so had more opportunity to contain or pick up nucleation catalysts.

BISMUTH

Although a major fraction of source *A* oxide coated bismuth particles supercooled 90° in dilatometric experiments only a minor fraction (presumed not to be oxide coated) supercooled the maximum of 90° on the microscope stage. However, it is significant that more than one-half of the particles made by breaking up source *B* bismuth rod in a mortar and pestle did not solidify unless supercooled about 90°. This difference in behavior of source *A* and *B* particles observed microscopically cannot be accounted for on the basis of different particle sizes for the *B* particles were significantly larger than the *A*. On the other hand, it seems probable that the difference in behavior of *A* particles in the dilatometric and microscopic experiments can be accounted for by the fact that many of the *A* particles observed on the stage had formed by coalescence of several small particles. As a result of this process, the proportion of particles containing nucleation catalysts was larger.

It is also significant that the temperature of solidification of bismuth particles that supercooled the maximum amount was not a function of the maximum temperature reached in the range of liquid stability when this was 10 to 100° greater than the melting point. This fact coupled with the lack of thermal history dependence in the dilatometric experiments proves that the thermal history effect observed in the solidification of large continuous bismuth samples is not an intrinsic property of bismuth and must be due to some extraneous effect.

The most important result of these experiments on bismuth is that the same maximum supercooling, 90°C, was obtained by two independent methods and with bismuth particles from two sources.

CRYSTAL GROWTH RATES

With the exception of selenium all of the droplets that had supercooled considerably solidified in a time too short to estimate, even to an order of magnitude,

[10] C. S. Smith, J. of Metals **1**, 204 (1949).

after a nucleus had formed. From these observations it can only be maintained that the linear growth rates were greater than 1 mm/sec. It seems likely that crystals with relatively more complex structures such as bismuth and germanium would grow into their melts much more slowly than cubic crystals but this could not be proved or disproved by these experiments.

The extremely short period of growth relative to the total time required for solidification of a particle proves that "nucleation" is the rate-determining step in the reaction. Here the term nucleation is used in the broad sense as referring to a period of very slow growth of small crystals relative to large for whatever reason.

In no case excepting selenium was it possible to form a "glass" from the metal particles by quenching directly from the melt. Liquid selenium particles cooled to room temperature without crystallization. However, the glass particles so obtained crystallized upon reheating to temperatures 100°C or greater.* By cooling rapidly to some given temperature below the melting point and reheating, it was established that selenium nuclei were formed when the supercooling was greater than 25–30°C.

KINETICS OF ISOTHERMAL SOLIDIFICATION

Up until now it has been tacitly assumed that it is possible to assign "solidification temperatures" to particular particles. Since solidification is a rate process some justification of this assumption is necessary. As has been pointed out already, qualitative observations of Turnbull[1] have indicated that the temperature coefficient of the solidification rate of small metal particles is so very great that the rate changes by several orders of magnitudes within a narrow temperature range. Thus, solidification will be observed to take place only within this narrow range for widely different rates of cooling from the melting temperature.

This was confirmed by some observations on gold and copper particles. Liquid gold particles were on several occasions held at temperatures 180 to 190° below their melting point for one to three hour periods without solidification. They did not then solidify until cooled 230° below the melting point that corresponds to $(\Delta T_-)_{max}$ reported in Table II. Similar results were obtained with copper particles chipped from the rod.

DISCUSSION

The basic phenomena involved in the solidification of small metal particles now seem clear. In general, particles 100 microns or less in size can be supercooled to an extent far in excess of that ever attained on large continuous samples. Only for iron has a ΔT_- been reported[11] for a large continuous sample that is of the same order of magnitude (258°) as $(\Delta T_-)_{max}$ (295°) measured for small particles. For other metals having cubic crystal structures it has rarely been possible to supercool large continuous samples of the liquid more than 5–10°.

In an aggregate of metal particles the solidification behavior of individuals is quite specific and persistent but a significant fraction supercools some maximum amount $(\Delta T_-)_{max}$ that is reproducible and characteristic of the particular metal. The ratio of $(\Delta T_-)_{max}$ to the absolute melting temperature T_0 is 0.18 ± 0.02 (see Table II) for the metals studied having cubic crystal structures excepting lead and aluminum. (In making this calculation for iron and manganese, it was assumed that the crystal modification in equilibrium with the liquid at atmospheric pressure was nucleated although solidification in both instances takes place in a temperature range where a different crystalline modification is stable.) The lowest temperature of solidification of small particles is varied only a small extent with widely different rates of cooling from the melt and with changes in the particle volumes amounting to factors of 10^9.

The interpretation of these phenomena is to be given in a succeeding paper by one of us.[8]

ACKNOWLEDGMENT

The authors are pleased to acknowledge the assistance of Mrs. Joyce Rowland on many of the experiments herein reported. They also thank Drs. Leroy Apker and W. C. Dunlap for supplying samples of some of the metals.

* This agrees with prior observations of Apker and Dickey (private communication) on the crystallization of amorphous selenium films.

[11] Bardenheur and Bleckman, Stahl u. Eisen **61**, 49 (1941).

Mass Spectrometric Study of Solids I. Preliminary Study of Sublimation Characteristics of Oxide Cathode Materials*

R. H. Plumlee and L. P. Smith
RCA Laboratories, Princeton, New Jersey
(Received March 6, 1950)

Mass spectrometric techniques have been utilized to study sublimation characteristics of materials used in electron tubes. Exploratory experiments and methods are described. Preliminary results concerning general processes induced by heat and by electron bombardment are discussed. Photographs are shown of mass spectrometer traces which demonstrate the effects described. Phenomena noted include (1) the inefficiency of electron bombardment for producing material transport from solids, (2) the thermal evaporation of alkali metal atoms and ions from bare nickel alloys, but no such evaporation from oxide cathodes formed on the same materials, (3) the complex thermal evaporation of oxide cathode constituents, (4) absence of positive ion evaporation from oxide cathodes, (5) the confirmation of mechanisms for negative ion formation in vacuum systems, and (6) the liberation of molecular oxygen from oxide cathodes which increases with electron emission drawn.

I. INTRODUCTION

ALTHOUGH considerable effort has been expended in the past on studies of structure-sensitive phenomena, the interpretation of the measurements produced has been inherently handicapped, for the most part, by a very deficient knowledge of the chemical constituents present in the materials studied. When the property under study, such as an electron emission process, is closely related to the functioning of the solid-vacuum interface as well as to the functioning of the interior of the solid and the solid-solid interfaces, the composition variable becomes quite complicated. The composition of surfaces is not only uncontrollable within the desired limits (because of diffusion exchanges with the interior of the solid, and because of absorption of foreign materials from the vacuum system), but it is also undeterminable by any currently known analytical technique. Surface analyses, ultimately of mono-molecular layers, made concomitantly with measurements of electronic properties, and without appreciably damaging the surface itself, would of course greatly facilitate interpretation of the measurements.

In developing an approach to the problem of surface analysis, mass spectrometric techniques are being applied to the study of materials of interest to the electronics industry. The results of these exploratory experiments described here are of interest in themselves, and, in addition, demonstrate the potentialities of a method applicable to a considerable number of problems in surface chemistry.

II. EXPERIMENTAL

The experiments were performed with a Nier-type spectrometer.[1] The tube was metal and glass with copper-gasket vacuum seals. The entire tube and liquid air trap were oven-baked at 450°C. A mercury diffusion pump was used. The spectrum was scanned by changing the ion energy. Recording was done by a carbon-paper printing mechanism from an obsolete facsimile receiver. The longest spectrum shown in succeeding figures represents less than five minutes of recording time. The ion accelerating potential was scanned linearly with time from high to low and vice versa at several fixed magnetic field values to cover the mass scale. The mass scale is thus inversely linear with ion energy, and the scales of some spectra are mirror images of others. The gaps in the strong peaks (e.g., Fig. 2A) are due to the recording arrangement whereby no printing was done during the scanning of a signal larger than the full ordinate scale (about 15 times the ordinate increment shown). Some of the experiments were recorded with an electron multiplier as detector. Most of the experiments were recorded with a conventional electrometer tube d.c. amplifier arrangement as detector. The ordinate increment shown for these traces represented usually an ion current of about 10^{-13} ampere. This has been translated into the estimated pressure sensitivity equivalent.

This sensitivity estimate was made from the measurement of the CO (mass 28) peak height and the associated ionization-gauge pressure reading during periods when the mass 28 peak constituted about 75 percent or more of the residual spectrum peak height. This estimate is necessarily rough because, besides uncertainties in the pressure measurement, it implies the same ionization cross section for all neutral particles. For many common gases, measured ionization efficiencies vary by less than a factor of ten[2] from that of CO. Although the ionization efficiencies of most of the species with which we are concerned have not been determined, this assumption of equal ionization cross sections is probably sufficiently accurate for present purposes. The necessity for maintaining the best possible vacuum discouraged the use of standard gases admitted for calibration purposes only. With solid samples being evaporated into the ionizing region, the sensitivity and optimum adjustment varied steadily so that frequent checks on the sensitivity were

* Presented in part at the meeting of the American Physical Society at Stanford University, December 30, 1949.

[1] A. O. Nier, Rev. Sci. Inst. **18**, 398 (1947).

[2] John T. Tate and P. T. Smith, Phys. Rev. **39**, 270 (1932).

essential. Using this calibrating method, fair agreement was found between the measured vapor pressure of nickel and literature values. For materials evaporated from the samples by the various methods employed, the measured pressures or vapor densities were also translated, by simple kinetic theory equations, into transport rates and into approximate times required for removal of mono-molecular layers.

The ion source most used was the double filament structure shown schematically in Fig. 1. It is essentially the usual gas analysis source with an added upper section for bombarding a heated sample. A similar design was used previously by Dart.[3] Bombardment may be done with positive ions (Kunsman source[4]) or with electrons. We have thus far used only electron bombardment. The sample bombarding region is separated from the ionizing region (analyzer) by a tungsten mesh. The structure therefore allows neutral particles evaporated from the sample to be analyzed in the usual gas analysis fashion, and charged particles from the sample region to be restrained, by suitable potentials, from entering the spectrometer analyzer; on the other hand, if so desired, charged particles from the sample may be accelerated through the spectrometer along with ions from the analyzing beam region, in which case two peaks for the same mass may appear. These peaks will be displaced from each other on the spectrum record by a distance corresponding to the potential difference between the two points of origin. Electron emission from

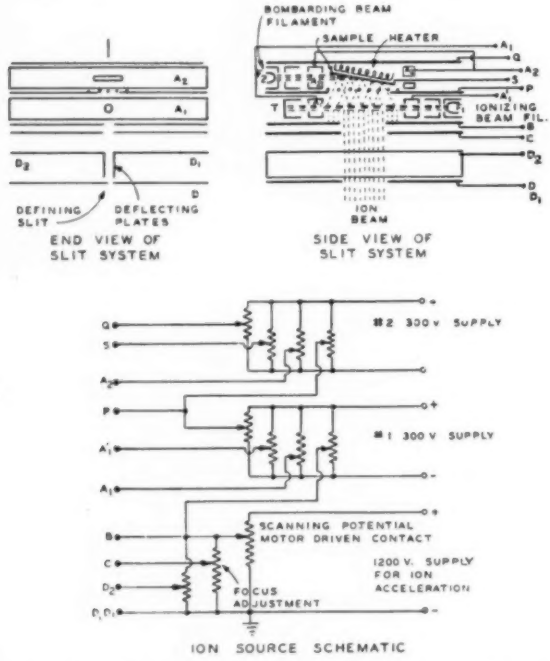


FIG. 1. Schematic drawing of an ion source and potential supplies used for electron bombardment study of solids.

[3] F. E. Dart, "The evaporation of zinc and zinc oxide under electron bombardment," doctoral thesis, Cornell University (1946); Phys. Rev. **78**, 761 (1950).

[4] C. H. Kunsman, Science **62**, 269 (1925).

the sample may be drawn to the surrounding box without traversing the mesh and without contributing to ionization in the analyzer region.

Both directly heated and indirectly heated samples were used, but mainly the latter. The sample was ordinarily in the form of a rectangular box (about $\frac{1}{2}$ in. long $\times\frac{1}{8}$ in. high $\times\frac{1}{8}$ in. wide) welded from a 0.002-in. sheet of the metal (or cathode base material) under study. A thermocouple junction of 0.0005 in. diameter Pt and Pt-Ir wires was welded to the sample, usually to the outside surface, centrally on the flat facing the spectrometer slit area. Oxide cathodes were made by spraying a carbonate mixture in a nitrocellulose-solvent vehicle over the box flat to be bombarded (the flat facing spectrometer slits). The sample was generally processed in the spectrometer. Processing was varied from experiment to experiment, but usually consisted of a brief operation, preliminary to bake-out, to check instrument performance, then a 450°C bake-out of several hours' duration with carbonate breakdown, the total pressure never rising above 10^{-3} mm Hg. After the carbonate breakdown, the liquid air was removed from the trap which was then torched empty of condensate, the liquid air was then replaced around the trap, and the bake-out continued for another hour or so. The residual pressure after bake-out was about 10^{-7} mm Hg with both source filaments operating and with the sample heated by its own heater to about 450°C. The source filaments were tantalum, coated on the emitting areas with tungsten powder and then thoria powder. They proved quite suitable for operations at pressures less than 10^{-5} mm Hg when no hydrogen was added to the residual gas of the system.

Figures 2A, 2B, 2C show typical binder decomposition spectra from an oxide cathode sample heated to 450°C before the bake-out. These materials, carbon-hydrogen-oxygen fragments of cellulose acetate and cellulose nitrate, were found to settle on all moderately cool elements of the ion tube and were not completely removed by the 450°C bake-out. They gradually distilled out of the ion source structure but could always be stirred up by moderately warming other portions of the ion tube.

Figure 3 shows the main residual spectrum at the end of the bake-out. The chief constituent of the residual gas was always CO. All residual peaks could be enhanced at any stage of operations by increasing the sample temperature, particularly in the case of oxide cathodes, indicating that an oxide cathode is never completely outgassed.

III. GENERAL ELECTRON BOMBARDMENT EFFECTS

Some cathode poisoning phenomena, well known to tube makers, have been attributed to dissociation or ionization of materials, adsorbed on grids or anodes, by electrons with energies from about 6 volts up to 25

volts.[5–7] A study of oxide decomposition using cathode poisoning as a method of detection was reported by Jacobs.[8] The technique of cleaning surfaces by electron bombardment has also been utilized.[9] Other experiments demonstrating material transport due to electron bombardment have also been described.[10] Dempster once employed electron bombardment of salts to generate ions for mass spectrometer studies.[11] Dart[3] found a small evaporation of Zn (and ZnO) from ZnO produced by electron bombardment.

It was expected, therefore, that electron bombardment might be the appropriate means of indiscriminantly stripping surface layers off solids for mass spectrum analysis; however, our electron bombardment studies of Ni, ZnO, Mo, Cu, and (BaSrCa) O surfaces have thus far found no general material transport effect that could definitely be separated from the purely thermal effect of the bombardment. No detectable evaporation of any sort was found to be produced by bombardment of these surfaces at low temperatures; however, at temperatures sufficiently high to produce a detectable thermal evaporation spectrum, addition of the bombardment, with the measured shell temperature fixed, enhanced the evaporation of non-metallic coatings only by slight amounts (less than a factor of ten, depending upon the bombardment). Electron bombardment energies up to 1400 ev and current densities up to 6 ma/cm^2 were used.

In no instance was an effect found which would indicate that surface layers of the bulk material of a solid are uniformly stripped away, at temperatures well below the sublimation point, by electron bombardment.

IV. SPECIFIC ELECTRON BOMBARDMENT EFFECTS

Although there may be numerous specific reactions between solids and low energy electrons, only one such was definitely established and one other probably established by these experiments. These are, respectively, the ejection of O^+ ions from (tank) hydrogen-fired molybdenum, and the ejection of Cl^+ ions from (tank) hydrogen-fired copper. The experimental arrangement by which the Cl^+ ions were obtained did not preclude the generation of these ions by negative ions (e.g., Cl^-) arising from the oxide cathode emission used for bombarding the copper. The ions were definitely formed at the metal surface and not in the gas phase.

Several times, indications were seen of O^+, OH^+, F^+, and Cl^+ liberation by bombardment of other metal surfaces, but the quantities were too small and transient for study. Circumstances suggest that they were all derived from adsorbed materials. Whether these materials were ionized on the surface before desorption, or ionized in the space adjacent to the surface was not operationally determinable.

The O^+ ions produced from the Mo surface had several unique properties, however, by which they differed from other ions in the spectrum. Figure 4 is a typical spectrum from that experiment. The ion source was the structure referred to earlier (Fig. 1) with the potentials $A_2-S=3$ v, $Q-A_2=0$ v, $S-P=10$ v, $P-B=2$ v, $A_1-B=1$ v, thereby transmitting positive ions from the sample region (superscript S peaks) as well as ions from the analyzer region (peaks without superscript S) to give a double spectrum. The displacement between peaks of the same mass should thus be expected to be

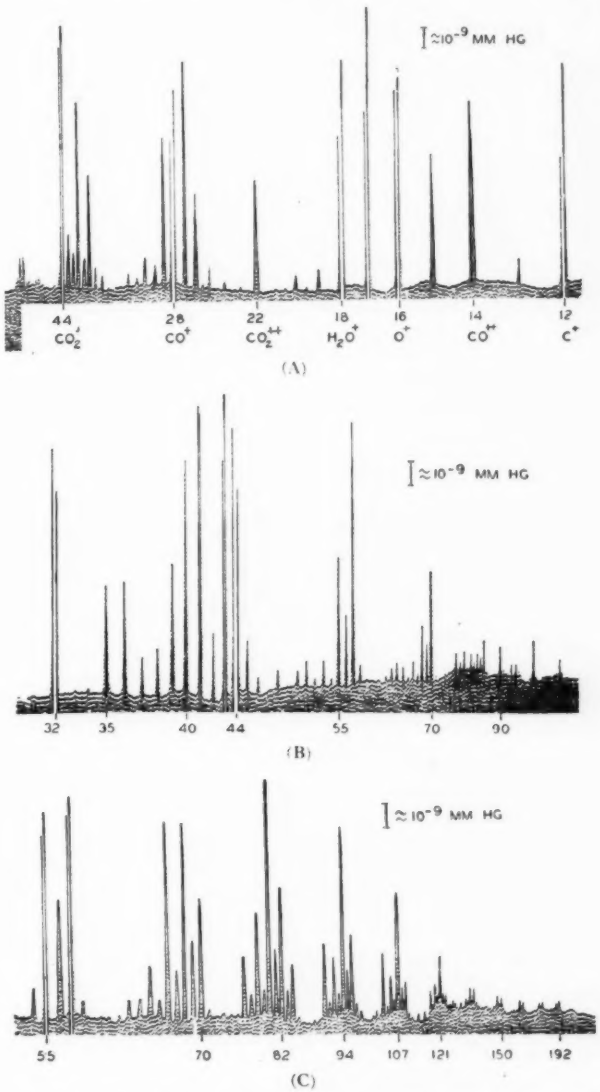


FIG. 2A, 2B, 2C, binder decomposition spectrum from oxide cathode sample at 450°C before bake-out.

[5] L. B. Headrick and E. A. Lederer, Phys. Rev. **50**, (A) 1094 (1936).
[6] Hamaker, Bruining, and Aten, Philips Res. Rep. **2**, 171 (1947).
[7] G. H. Metson, The Post Office Electrical Engineers Journal **41**, 204 (1949) (Part 4); Proc. Phys. Soc. (London) **62B**, 589 (1949).
[8] Harold Jacobs, J. App. Phys. **17**, 596 (1946).
[9] J. T. Randall, *The Diffraction of X-Rays and Electrons by Amorphous Solids, Liquids, and Gases* (Chapman and Hall, Ltd., London, 1934), p. 236.
[10] L. Jacob, Nature **157**, 586 (1946). G. I. Pokrovskii and S. T. Sinitsyn, J. Exp. Theor. Phys. (USSR) **8**, 1174 (1938). Chem. Abs. **33**, 7660[3] (1939). J. G. Trump and R. J. Van der Graaff, J. App. Phys. **18**, 327 (1947). I. Filosofo and A. Rostagni, Phys. Rev. **75**, 1269 (1949).
[11] A. J. Dempster, Phys. Rev. **11**, 316 (1918).

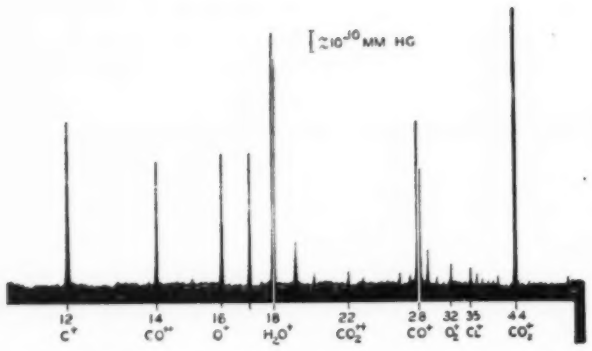


FIG. 3. Main portion of the residual spectrum after bake-out (and carbonate breakdown); sample temperature 450°C.

about 11 volts [the sum of $(S-P)$ and $\frac{1}{2}(P-B)$] corresponding to the potential difference in the two points of origin. Measurement of this displacement gives reasonable agreement for masses 18, 19, 23, 28, and 39, but not for the mass 16 peaks, for which the displacement is somewhat larger. Operation with a reversal of the $S-P$ field suppressed all the S peaks, less than 2 volts being required to suppress 18^s, 23^s, 28^s, 39^s, but about 11 volts being required to suppress 16^s completely. The excess kinetic energy of the O^+ ions (16^s peak) is the first property unique to those ions.

The second property of the 16^s peak not common to other S peaks was its decay under bombardment. The decay was roughly 50 percent in two minutes under bombardment of 0.9 ma per cm² with 300-v electrons and was independent of temperature. After an interlude of no bombardment, the peak could again be obtained at the original height upon resumption of bombardment. Even bombardment with 1000-v electrons provided a period of partial recovery from bombardment with 300-v electrons.

The third property of the 16^s peak was its temperature dependence. It could be obtained at temperatures above 300°C, and increased with temperature. It disappeared, when operations were suspended overnight, and could be obtained again only after a brief heating of the sample to about 900°C, a circumstance which suggests reaction of the Mo with residual gas of the vacuum system, or perhaps diffusion of some material from within the metal to give the surface coating required. The 16^s peak was produced with electrons of energy ranging from 40 volts to 1400, the highest used, with a broad maximum around 140 v. Below 140 v, the efficiency dropped gradually to extinction at 40 v. Since the remaining peaks (excepting the alkali metal peaks 23, 23^s, 39, 39^s, 41^s) showed almost no temperature dependence, they were concluded to be characteristic of the residual gas of the system. Sodium and potassium evaporated thermally from the sample as both atoms and ions. Upon application of bombardment, the increase in the 23^s and 39^s peaks due to electron impact ionization of metal atoms evaporating was small and the ratio of, for example, the regular 23 peak to this increase was not far from the ratios of $18/18^s$ or $28/28^s$ peaks. This established the relative transmission efficiencies for ions produced from the sublimate by the two beams. The $16/16^s$ ratio showed no relationship to this transmission ratio. Thus the 16^s ions must have originated at the surface and not in the vapor phase ionization of neutral particles evaporated from the surface. The excess kinetic energy indicates a dissociation from some excited state produced by bombardment of the surface layer.

V. THERMAL EVAPORATION FROM METALS

Studies of thermal evaporation from several† bare metals (cathode nickel alloys, and one molybdenum sample) showed, at moderate temperatures, copious evolution of alkali metal ions, and, at higher temperatures, evolution of alkali metal atoms as well as ions. At the higher temperatures, nickel alloys produced the expected evaporation of nickel atoms, and a rather profuse spectrum of carbon-hydrogen compounds which obscured the evaporation of minor constituents such as Mg, Mn, Ti, and Si. Molybdenum showed less carbon-hydrogen compounds. All samples had previously been fired in hydrogen.

The alkali metal evaporation was detected roughly as follows from nickel alloys: K^+ above 500°C, Na^+ above 600°C, K^0 and Na^0 above 850°C, the evaporation of all four species increasing with temperature, and the ratio of atoms to ions rising sharply with temperature. It may be significant that oxide cathodes made from the same nickel alloys showed no traces of alkali metal evaporation either as atoms or ions, at temperatures as high as 1100°C.

Referring again to Fig. 4, the 23^s, 39^s, and 41^s peaks are due to $^{23}Na^+$, $^{39}K^+$, and $^{41}K^+$ evaporation. These peaks appeared with both electron beams turned off, and showed temperature dependencies different from those of the regular 23, 39, and 41 peaks obtained with a retarding field ($P-S=5$ v) to prevent ions from the sample region entering the analyzer, and with only the ionizing electron beam turned on. On the other hand, the peaks for masses 12, 16, 17, 18, 18^s, 19, 19^s, 28, 28^s, and 44 were almost independent of sample temperature, and were thus identified as characteristic of the residual gas of the vacuum system. The total pressure of the system was less than 10^{-7} mm Hg according to ionization gauge readings.

VI. THERMAL EVAPORATION OF NEUTRAL PARTICLES FROM OXIDE CATHODES

The thermal evaporation spectrum from oxide cathodes depended very much on previous thermal history of the samples. In the early stages of the heat treatment (after carbonate breakdown and a "hot shot" at 1100°C to simulate activation) the evaporation from (BaSrCa) O cathodes was largely that of barium compounds, some

† A comparable thermal study of Cu was not made.

Ba atoms, and less Sr. The barium compounds evaporating were identified as BaO, and $BaCl_2$ (Fig. 5). (See Section XI.) The early evaporation, detected at temperatures as low as 600°C decayed at roughly 1 percent per minute and could be maintained at the minimum detectable level (about 10^{-10} mm Hg pressure in the ion source region) only by steadily increasing the sample temperature. During several hours of this procedure, the barium chloride evaporation rate declined to values small compared with the BaO evaporation. However, even though samples were held at temperatures around 800°C, for as long as 15 hours, the barium chloride could subsequently be made to register at sample temperatures of 1000°C to 1100°C. It is noted that the barium isotopes contribute fine structure to the peaks, and also that the Ba^+ peaks consist of contributions from ionization of both Ba and Ba compounds in the analyzer. It is also to be remembered that the evaporation is entirely uncharged, the ions being formed in the analyzer and not at the sample.

Prolonged operations with sample temperatures at 950°C±100°C developed, after several hours, a much slower decay rate for the Ba, BaO evaporation, of the order of a few percent per hour. The Ba^+/BaO^+ ratio was never larger than about 2/1 and declined slowly throughout. The ratio was definitely lower for the one platinum based cathode examined (Fig. 6) than for any of the nickel based cathodes formed from alloys containing medium or low quantities of reducing agents. The lowest ratio Ba^+/BaO^+ observed was about 0.4. The minimum peak ratio, Ba^+/BaO^+, is to be derived of course when no Ba evaporates, and is determined by the relative probabilities of the various decomposition processes with which ions are formed by the electron beam of the analyzer. Figure 7 shows even an appreciable Ba^{++} peak resulting from these processes.

The fact that some of the spectra shown were taken with the samples under bombardment is of no significance since the bombardment seemed merely to be an additional source of heat applied to the coating. Potentials were applied so that vapor-phase ions from the bombarding beam were not transmitted through the analyzer. In all cases, the absolute peak heights depended upon the past thermal history of the sample.

In general, the cathodes on nickel bases showed, in addition to the larger Ba/BaO ratio in the evaporation, appreciable Sr evaporation which was often nearly as large as the Ba evaporation (Fig. 7). At higher temperatures, a small Ca evaporation was detected also. The Sr evaporation decayed much less rapidly than the Ba and BaO evaporation.

Only traces of SrO evaporation were found, even at 1100°C, the highest temperature used. At this temperature, the literature[12] value of the vapor pressure of pure SrO is about 10^{-9} mm Hg, about ten times the minimum detectable by the instrument in some experiments. Considering the fact that high temperature effects were examined only toward the end of the experiments when the sensitivity may have diminished because of nonconducting coatings evaporated upon slit edges, the discrepancy between the literature value and the observed evaporation is not excessive. All samples were triple oxides, so that depletion loss of the more volatile BaO and SrO from outer layers of the coating would eventually leave an outer layer consisting very largely of CaO. At any stage, the vapor pressure of each constituent would be approximately equal to the product of its mole fraction and the vapor pressure of the pure material.

The cathodes examined were not what would be considered "well-formed," good emitters. This may have been due to contamination from the ion tube, or to improper activation. Emission values were usually about 25 ma d.c. per square centimeter for a cathode at 800°C.

[12] John P. Blewett, J. App. Phys. **10**, 668 (1939). A. Claassen and C. F. Veenemans, Zeits. f. Physik **80**, 342 (1933).

VII. POSITIVE ION EMISSION FROM OXIDE CATHODES

No positive ion emission was found from oxide cathodes at temperatures as high as 1100°C. The instrument was capable of detecting positive ion evaporation smaller than 10^{-10} amp. per cm^2. This does not conflict, of course, with literature references to positive ion emission from (BaSrCa)O coatings on tungsten at much higher temperatures.

VIII. THE NEGATIVE ION SPECTRUM

Examination of the negative ion spectrum agreed in general with the observations summarized by Sloane and Watt,[13] namely, that three principal mechanisms may be responsible for negative ion formation in partially evacuated systems, (1) thermal evaporation from surfaces, (2) conversion‡ of positive ions striking a surface to negative ions, which subsequently escape, often carrying part of the kinetic energy of the positive

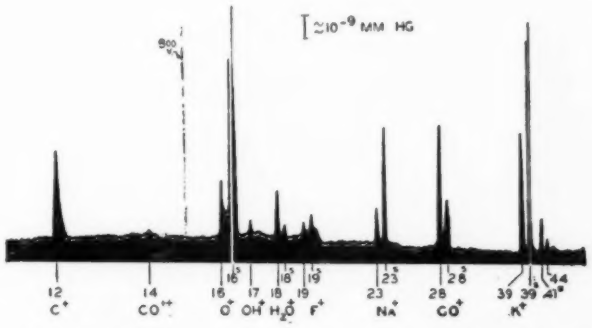


FIG. 4. Thermal evaporation of alkali ions and atoms from (H_2 fired) molybdenum at 1100°C, and ejection of O^+ ions by electron bombardment (4 ma/cm^2, 300 v) of molybdenum surface.

[13] R. H. Sloane and C. S. Watt, Proc. Phys. Soc. (London) **61**, 217 (1948).

‡ F. L. Arnot, Proc. Roy. Soc. (London) **158A**, 137 (1937) gives yield values, N_-/N_+, for this process ranging from 10^{-3} to 10^{-5} with 180-v positive ions.

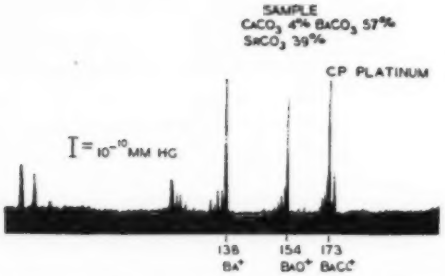


FIG. 5. Oxide cathode initial thermal evaporation at 800°C, all peak heights slightly exaggerated because of electron bombardment (4 ma/cm², 300 v).

ions, and (3) positive ion impact causing the dislodging of absorbed materials¶ as negative ions, which may also carry away some of the kinetic energy of the positive ions. These latter two processes were not distinguishable in the experiments here reported.

The operation demonstrated by Fig. 8 was performed with the source structure of Fig. 1, but with both electron beams turned off. Accordingly, with all source potentials negative with respect to ground and the ion collector, and with $P-S=100$ v, $B-P=2$ v, negative ions and electrons from the sample were accelerated through the spectrometer. In the spectrum, the displacement between peaks of the same mass number was found to be the voltage difference between the mesh P and sample S, and the non-superscript S peaks followed the potential of P. Therefore, these peaks were concluded to originate at the mesh while the S peaks were concluded to originate at the sample. With the field between S and P reversed so electrons and negative ions could not enter the analyzer region, no peaks were obtained. With $P-S=5$ v so electrons and negative ions were drawn from the sample to the mesh with low energy, then through the analyzer with high energy (several hundred volts), the non-superscript (mesh) peaks were again produced. This suggests that they were independent of the bombardment of the mesh by electrons or negative ions. Use of a magnet to deflect electrons (but not ions appreciably), so that the electrons did not accompany the ions through the analyzer, cut out all except the Cl^- ions. The interpretation is then that the large current of electrons in transit through the analyzer ionizes the residual gas to form positive ions which then follow the reverse path of the electrons striking slit edges, the mesh P, and finally the sample. On these surfaces, the incident positive ions generate the negative ions which are affected by the presence, in the source region, of the auxiliary magnetic field.

The largest negative ion peaks were those of the chlorine isotopes (Fig. 8) which were both temperature

¶ R. H. Sloane and R. Press, Proc. Roy. Soc. (London) **168A**, 284 (1938). Their results, given for CO^-, Hg^+, CO^+, Hg^{++}, indicate current density yield ratios, N_-/N_+, for this process larger than 10^{-5}, the uncertainty arising from the measurement of the positive ion current. An unknown degree of surface coverage by adsorbed CO is also involved in this yield factor so the yield per impact per CO site is indeterminable.

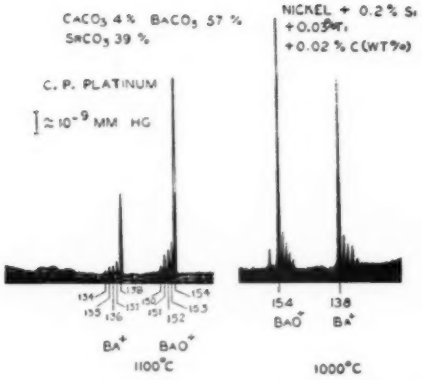


FIG. 6. Comparison of oxide cathode evaporation from platinum and nickel bases.

and pressure dependent, and were concluded to be largely of the thermal evaporation variety. They were found evaporated from an oxide cathode at temperatures as low as 600°C. The Cl^- peaks were much more sensitive to torching of the ion tube at various points than were the Cl^+ peaks formed in the usual operation of the analyzer. This observation suggests that the cathode collects neutral chlorine compounds or atoms and ejects the chlorine as negative ions. It also implies some storage mechanism whereby the Cl^- ejection need not be dependent upon the instantaneous vapor phase concentration of chlorine compounds as is the Cl^+ measured, but rather upon the concentration of chlorine accumulated in the cathode.

Masses 26^- and 14^- (Fig. 8, the superscript s peaks) were also generated at the cathode surface, but probably by one of the positive ion mechanisms, since there was little temperature dependence.

Mass 14^- was generated also, along with 12^-, at the unheated mesh partitioning the sample region from the analyzer section of the ion source (Fig. 1), by positive ion bombardment.

Trace quantities of still other negative ion masses were also found, such masses as 42, 68, 75, 130, but they were too small and erratic in appearance for study. They were apparently generated by a positive ion bombardment mechanism.

There was no visible correlation between the negative ion spectrum obtained as above and the usual residual gas spectrum (compare Fig. 3) obtainable a few minutes previously or subsequently (the change-over time from negative ion search to positive ion search was one to two minutes). In view of the variety of species inhabiting the vacuum system, any attempt to identify the composition of ions less obvious than C^-, $^{35}Cl^-$, $^{37}Cl^-$, seemed entirely speculative.

IX. EMISSION EFFECTS

Only one definite effect of electron emission from an oxide cathode was found reflected in the mass spectrum. This was an increase in the molecular oxygen peak

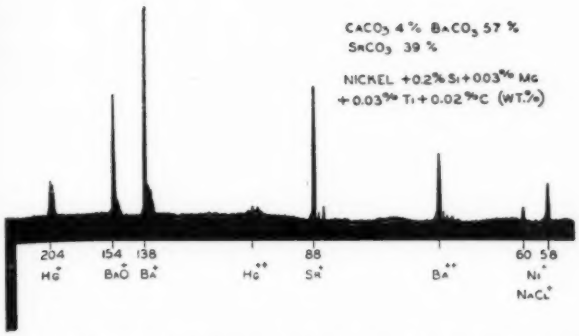


FIG. 7. Oxide cathode evaporation spectrum showing appreciable Sr evaporation at 1000°C.

(mass 32, Figs. 9, 10) as a function of emission drawn from the sample. The emission was drawn to the box surrounding the sample so it did not traverse the mesh P and contribute to the ionization in the analyzer. The bombarding filament was of course turned off during the experiment.

The total gas pressure was less than 10^{-5} mm Hg of which the O_2 partial pressure was probably never more than 1 percent. Examination of the negative ion spectrum showed no evolution of O^- or O_2^-. The evidence thus supports the conclusion that the oxygen left the cathode surface as O_2 molecules. The ratio, N_{O_2}/N_e, of oxygen molecules to electrons evaporated from the sample was estimated at about 10^{-5}.

X. CATHODE BREAKDOWN

The most significant fact of the breakdown was that it was never completed—each of the usual components (Fig. 11) seems to be evolved at a dwindling rate throughout the cathode life. The CO evolution, after the bulk of the decomposition was completed, was the most intense and by far the slowest to decay. The CO_2 and H_2O evolution usually assumed relatively small rates after several hours of sample operation at one particular temperature, only to be renewed at higher, slowly decaying, levels when sample operation was carried to a higher temperature. For nickel alloys and oxide coated samples, CO pressures less than an estimated 10^{-7} mm were never observed; for the Mo sample the CO pressure was about tenfold lower.

XI. DISCUSSION OF RESULTS

A. Electron Bombardment Experiments

The bombardment experiments indicate first that, if there is a general vaporization or dissociation characteristically produced by electron bombardment of solids, it is too small at low temperatures for detection by ordinary mass spectrometric methods. If, as has been proposed,[3] bombardment transport is dependent upon high temperatures to permit escape of excited fragments before de-excitation by other means, it may not be separable from ordinary thermally activated reactions and evaporations occurring at the same temperatures. All cases of simple vaporization or dissociation induced by electron bombardment must be carefully examined to be certain that they are not derived thermally from the temperature rise of the surface. Very few experiments reported in the literature demonstrating mass transfer induced by electron impact are above suspicion in this respect.

The second aspect of the electron bombardment experiments is the evidence that surface layers may undergo specific reactions. These may be expected to vary in efficiency from values approaching those of impact processes in free molecules down to zero, depending upon the particular nature of the layer under bombardment. Various properties of both the surface material and substrate, in addition to simple bond energies, may be critical.

Since these experiments are exploratory in nature and inherently lacking in precision, only a semiquantitative estimate of the results can be presented. The minimum detectable yield ratio, N_0/N_e, of neutral particles removed to bombarding electrons incident at the target, was about 10^{-6}. The minimum detectable yield ratio for positive ions ejected by electron bombardment was $N_+/N_e \cong 10^{-8}$ (because no subsequent ionization step was necessary).

In the case of O^+ ejection from the Mo surface, the 50 percent decay time (e.g., 2 minutes under 1 ma/cm², 300-ev bombardment) was temperature independent, but the peak height increased with temperature. The current density yield ratio, N_{O^+}/N_e, was about 10^{-6} at 1100°C and 10^{-8} at 300°C for 300-v electrons. Because of the constant decay time, it was concluded that the yield variation reflected a temperature dependence of the oxygen supply rather than a temperature dependence of the removal cross section. Assuming that the removable oxygen was adsorbed on the Mo surface, either as atoms or in some molecular form, one O atom at most per Mo site, and that half the supply of O sites was removed in 2 minutes of bombardment (the recovery time was about 5 minutes) and assuming a (100) Mo plane surface, 3.14A unit cell length,[14] the bombard-

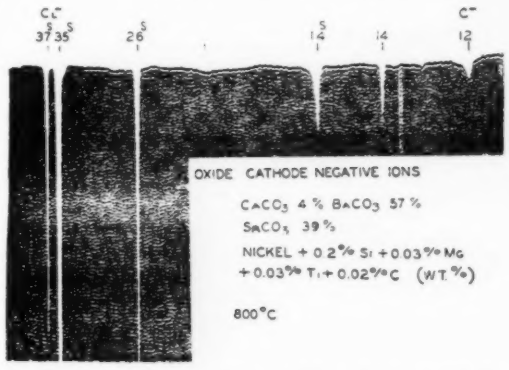


FIG. 8. Negative ions obtained from oxide cathode sample (^{s} peaks) and from unheated tungsten mesh (mass peaks 12, 14).

[14] R. W. G. Wyckoff, *The Structure of Crystals* (The Chemical Catalog Company, Inc., New York, 1931), second edition.

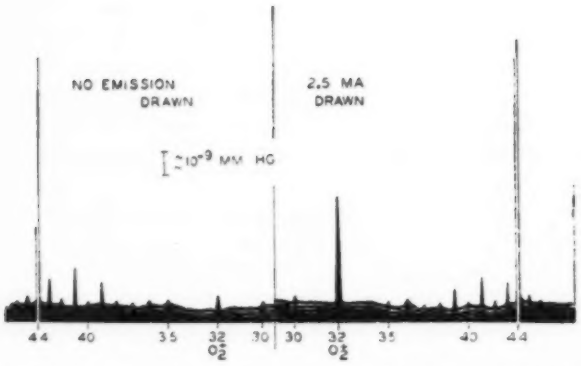


FIG. 9. Gas spectrum in the presence of oxide cathode sample showing increase in O_2^+ peak height with electron emission drawn from sample.

ment required for removing the adsorbed layer corresponded to about 10^3 electron impacts per surface atom site. Over-all yields of 10^{-6} and 10^{-8} thus corresponded to O site concentrations of one per 10^3 and 10^5 Mo sites, respectively.

Scant observations of the Cl^+ removal from Cu by bombardment with emission from an oxide cathode gave a single estimate of $N_{Cl^+}/N \cong 10^{-7}$. No decay was observed. If the Cl^+ resulted from conversion of Cl^- ions in the incident beam, the yield must have been near unity. This latter mechanism is considered rather unlikely in view of the fact that the Cl^+ ions were formed with essentially zero energy at the copper surface for bombarding beam energies up to several hundred volts. Had they been related to incident Cl^- ions in the beam, they would probably have carried away some of the kinetic energy of the incident ions.

It should perhaps be emphasized again that although yields as low as one neutral for 10^6 bombarding electrons or one ion for 10^8 electrons could have been detected, no such erosion of the bulk material of targets [e.g., Ni atoms from nickel sheet or Ba from (BaSrCa)O] was found, and the only removal observed was that of ions, formed from foreign materials adsorbed in some fashion on the target surface. Compared with reported positive ion (150 ev energy) sputtering yields[15] near 10^{-1} (one atom to ten ions) general electron sputtering yields are definitely trivial.

An application of these yield figures to the questions of target erosion and cathode poisoning is of interest. For the practical condition of electron emission (300 ma d.c./cm²) being drawn from a cathode to produce a general anode erosion yield of 10^{-6} (which could have been detected experimentally but which was not found), a maximum transport of 2×10^{12} neutral particles per second per cm² would be achieved.|| If this were all deposited on the cathode, one atom site in about 350 would be covered in one second, assuming a smooth (100) BaO plane of 5.50A[14] unit cell length for the cathode surface. In 1000 hours of cathode life, the anode erosion (e.g., copper) would amount to about 10^{-4} inch maximum. Thus anode erosion by this process would be insignificant, and sudden poisoning by the anode material would be unlikely.

Cathode poisoning of the low voltage variety[5-8] is usually associated with the drawing of electron emission to previously unbombarded tube elements and has been attributed to desorption, dissociation, or ionization of materials adsorbed on the target. Often beam energies too low to produce ionization are effective. The poisoning is usually observed to proceed to completion in less than one minute. A transport probably adequate for poisoning, about 0.1 monolayer cathode coverage per second, could be produced by the 10^{-3} efficiency observed for O^+ removal from Mo with space-charge-limited emission (30 ma/cm², 10 v, 0.5 mm spacing) drawn to a well-covered anode. This efficiency, computed as an effective collision cross section, is about 1 percent of peak values found for ionization of simple gases, hence higher values could occur. However, the gross removal efficiency at the critical anode potential may be indefinitely smaller than the peak value, because, depending upon cathode sensitivity to the poisoning agent, thermal energy distributions may determine the apparent threshold.

For pulsed emission operations (e.g., 30 amp./cm², 1 percent duty) the transport per second, aside from thermal effects, would be the same as for the 300 ma/cm² d.c. case; and for a 10^{-6} yield, in a single 100-micro-

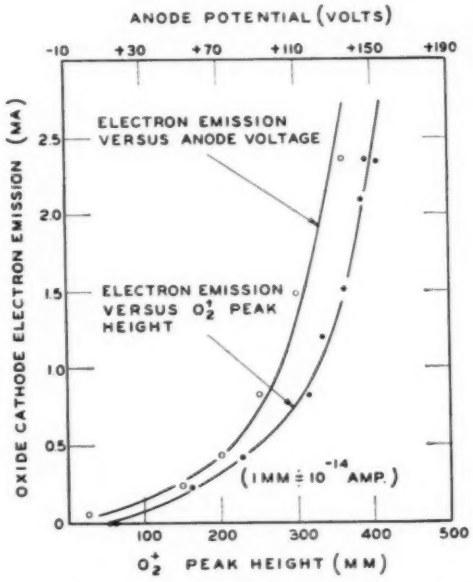

ANODE POTENTIAL (VOLTS)
-10 +30 +70 +110 +150 +190
OXIDE CATHODE ELECTRON EMISSION (MA)
0 0.5 1.0 1.5 2.0 2.5
ELECTRON EMISSION VERSUS ANODE VOLTAGE
ELECTRON EMISSION VERSUS O_2^+ PEAK HEIGHT
(1 MM $\doteq 10^{-14}$ AMP.)
100 200 300 400 500
O_2^+ PEAK HEIGHT (MM)

FIG. 10. Plot of electron emission drawn from oxide cathode (about 0.1 cm² coating area) *versus* O_2^+ peak height.

[15] K. H. Kingdon and Irving Langmuir, Phys. Rev. **22**, 148 (1923).

|| For comparison, the intervening gas space maintained at a pressure of 10^{-6} mm Hg would contribute about 10^{14} incident gas particles per second per cm² of cathode, and would yield also about 10^{12} positive ions per second per mm of electron beam path (assuming 100-v electrons and a gas having the molecular weight and ionization cross section of CO). There would be only about 10^{10} molecules per cm³ of gas space.

second pulse, 1 percent of the d.c. transport for one second could accrue, 3×10^{-5} monolayer maximum. As indicated previously, about the same transport of positive ions might be derived from the gas space. Deactivation in a single 100 microsecond pulse as a result of reactions initiated by electron transit would thus hardly seem feasible. Higher yield electron bombardment processes, such as the ejection of adsorbed materials from well-covered anodes would appear essential to produce emission decay in this interval.

B. Evaporation Experiments

The evaporation of alkali metal impurity ions from other metals has been widely studied.[16] The re-evaporation of alkali metals as atoms and ions from heated filaments on which they had previously been deposited is also well known.[17] From these data it would be expected that, where impurity alkali ions are found evaporating from heated surfaces, atoms might be observed evaporating also. This has been reported[18] for Kunsman sources; however, we have found no reference reporting that impurity alkali atoms were detected evaporating from metals. The alkali metal atoms may conceivably have quite an important role in the reactions postulated to occur in oxide cathodes, since it was found that neither the alkali ions nor atoms evaporate from (BaSrCa)O coatings on metals which, when bare, gave a large evolution of alkali ions and atoms.

The evaporation of neutral entities from oxide cathodes was somewhat complicated. As anticipated from the accepted picture of the evaporation, the more volatile barium compounds were lost quite rapidly in the initial heat treatment and then much more slowly at a gradually diminishing rate. This agrees with the suggestion of Claassen and Veenemans[12] that the outer layers of individual crystals of (BaSrCa)O are quickly depleted of BaO and the rate shortly becomes limited by the diffusion rate.

The evidence of $BaCl_2$ evaporation from cathodes is based on the identification of peak groups (Fig. 5) as isotopic $BaCl^+$ (isotopic in both Ba and Cl). This identification is made with some hesitation in that no trace of $BaCl_2^+$ was found in the spectrum. These peaks were originally identified as BaO_2^+, $Ba(OH)_2^+$ from

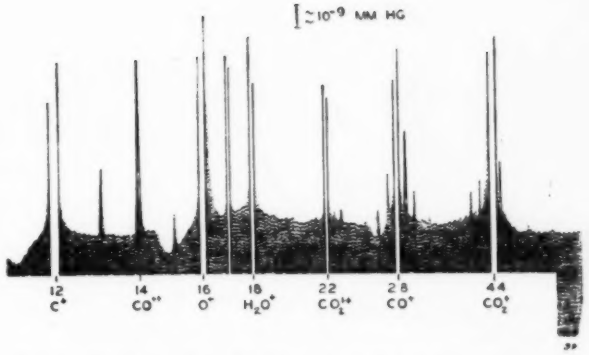


FIG. 11. Main portion of oxide cathode breakdown spectrum.

BaO_2 and $Ba(OH)_2$, respectively. Duplication of the result on several successive samples, together with the constancy of the isotopic (chlorine) ratios and better fit on the mass scale calibration favors the $BaCl^+$ identification. A $BaCl_2$ evaporation was reported by Hamaker, Bruining, and Aten.[6]

The Ba evaporation was expected, on the basis of the "free barium" picture and the mechanism of continuous core reduction of BaO, to be the most prominent feature of the cathode evaporation spectrum. The fact that it was always approximately equal to the BaO evaporation suggests it is not independent of the BaO evaporation, hence that it is perhaps limited by the rate of BaO reduction which in turn is limited by the rate of BaO diffusion to the reducing agent (and not by diffusion of reducing agent to the BaO). The presence of Sr in the sublimate, without SrO, argues against the possibility that the bulk of the Ba was oxidized by the residual gas as fast as it was generated and that the Ba^+ in the spectrum was all derived from ionization of BaO. The evaporation phenomena suggest that a considerable amount of information concerning the reaction mechanism could possibly be derived from a thorough study of evaporation rates and ratios and their decay characteristics.

XII. ACKNOWLEDGMENTS

Valuable assistance was derived from the suggestions of other members of the Physical Research Group, particularly Dr. D. O. North. Acknowledgment is made also of important contributions to the instrumentation by Dr. A. T. Forrester, Mr. W. B. Whalley, and Mr. J. F. Williamson, Jr., former members of the Laboratories staff. The recording system was designed by Dr. R. E Shrader.

[16] A. L. Reimann, *Thermionic Emission* (Chapman and Hall, Ltd., London, 1934).
[17] J. B. Taylor and I. Langmuir, Phys. Rev. **44**, 423 (1933).
[18] A. M. Tyndall and C. F. Powell, Proc. Roy. Soc. (London) **136A**, 145 (1932).

The Static Electrification of Dust Particles on Dispersion into a Cloud

W. B. KUNKEL*
University of California, Berkeley, California
(Received February 17, 1950)

An extensive study of the charge and size distribution of particles from 0.5 to 30 microns radius in dust clouds of diverse exemplary types dispersed in air under a variety of conditions ranging from blowing with minimum of turbulence to conditions of violent and maximum impact with various types of surfaces was made using adaptations of the Hopper and Laby modification of the oil drop experiment. It was observed that all dusts, including homogeneous dusts with no impacts on solid surfaces, were charged. In homogeneous systems the charges of opposite sign were equal, no *net* charge resulting, the number of particles of similar size with opposite charges being about the same. The magnitude of charges increased somewhat less rapidly than the surface of the particle. There was no correlation between size and sign of charge. There is strong evidence that charging occurs on separation of the contacts between particles in the dispersion of the cloud. Humidity did not affect the charging. Studies of heterogeneous systems making contact with solid walls of different composition from the powder gave consistent asymmetry of charge of varying degrees depending on the proportion of particles striking the surfaces relative to those just separated.

ALTHOUGH frictional electrification is the oldest manifestation of electricity known to man, it still remains today quite obscure as to the mechanisms active.[1] Not only are its mechanisms little known, but most of the data obtained are not reproducible and, in general, appear to be highly inconsistent and to a large extent even contradictory.

Most studies of frictional electrification antedate 1933 and involved studies using macroscopic systems. Work at this time, though stimulated by industrial processes affected by such electrification, appears to have been abandoned, primarily because with the many uncontrolled variables, the existing techniques were inadequate to yield consistent results. This is not surprising since most studies concerned themselves with net charges produced using particles of different sizes on carefully treated surfaces with controlled surroundings, temperature and pressure. As will appear evident later, the net charge analysis concealed some very important phenomena whose existence is useful in interpretation. The conclusions drawn were that charging depends on properties of the surfaces and involved the nature of the contact. Very small changes in these factors sometimes entailed large variations in the observed effects. The work of D. Debeau[2] in this laboratory showed the reproducible influence of absorbed gas films on surfaces of Ni, quartz, and rock salt, as well as water vapor exhaled from the walls. Recent studies of Peterson[3] showed similar changes of quartz surfaces by condensation of impurities exhaled from metal surfaces after evaporation.

That perhaps more information might be gleaned by studies of the individual charges on large numbers of microscopic particles with large surfaces relative to their masses, seems to have occurred to numerous workers. Thus, H. Sachsse,[4] in 1932, determined the charges on dust particles using the Millikan oil drop method. Similarly, it occurred to L. B. Loeb quite independently, for in 1933 he assigned to S. Chapman[5] the analysis of spray electrified particles by the same method. The method is too slow and tedious to obtain good statistics. Nevertheless, Sachsse found the charges to depend on the material as well as the methods by which dust clouds were generated. In most cases, positive and negative dust particles were present in equal numbers with only a small number of neutral particles. Using compressed air for dispersal, the average charge was found to be roughly proportional to the particle surface, but in most other cases, it seemed to be proportional to the radius. The observed charges ranged from 50 to 2500 electrons. In general, Chapman's results with atomized liquid particles confirmed these findings. The charges of both signs were present with few uncharged droplets for materials of high dielectric constant. A pure hydrocarbon gave 25 percent uncharged particles and low charges of both signs on the rest. The magnitude of charge was nearly proportional to drop radius and charges of up to 800 electrons were observed for the largest drops of 10-microns diameter. In 1937, N. Rosenblum[6] studied the behavior of charged dusts in an electrical field of alternating polarity. The amplitude of the photographed motion of the particles gave the charge, while the vertical motion gave the mass. The findings, in general, confirmed those of Sachsse.

With the contemporary intensified study of the solid state, the nature of solid surfaces thus derived may ultimately serve as a guide to experimental studies that will reveal much of the nature of the charge transfer at work. The improved modern techniques now available should also facilitate study, using microscopic systems so as to lead to consistent results such as Debeau observed. Currently, however, the opportunity offered since 1938 to study the charges and sizes of individual

* This work was supported by ONR Contract.
[1] L. B. Loeb, Science **102**, 573 (1945).
[2] D. Debeau, Phys. Rev. **66**, 9 (1944).
[3] J. W. Peterson, Phys. Rev. **78**, 91 (1950).
[4] H. Sachsse, Ann. d. Physik **14**, 396 (1932).
[5] S. Chapman, Physics **5**, 150 (1934).
[6] N. Rosenblum, Tech. Phys. USSR **4**, 564 (1937).

particles of a cloud of microscopic particles in a rapid fashion with good statistics seemed too good to pass over. Thus, as soon as the work of Hopper and Laby[7] on the evaluation of the electron using oil drops falling vertically in a horizontal electrical field and observed photographically by dark field illumination periodically interrupted by a rotating shutter was published, L. B. Loeb put students to the task of adapting this device for the study of dust clouds. The first attempt was interrupted by the war. In 1946, J. W. Hansen, an entomologist interested in the problem of insecticidal dusting, who had previously studied the electrification of such dusts using the older techniques, volunteered to set up the devices. Through his painstaking and skillful technique, the device was made to operate successfully and preliminary results were obtained.[8] Subsequently, the writer undertook to improve the device and with the improved techniques, which are described in a separate joint article by Hansen and the writer, the data here presented were obtained. As will be noted, the results obtained throw a considerable light on the mechanisms at work. There are still too many uncontrolled variables, as yet, to yield a complete picture of the processes at work, but certain mechanisms are definitely ruled out and other points clarified. The method has, of course, its limitations, but much more work with controlled atmospheres, as well as different systems, is still possible within the limitations encountered. Owing to the time involved in making the essential modifications for such work, this was not undertaken at this time. At a later date, it is possible that such studies will be undertaken.

EXPERIMENTAL TECHNIQUE AND LIMITATIONS

As indicated, the details of the apparatus and method of operation are elsewhere presented.[9] This paper will thus only concern itself with those features of the investigation affecting the results to be reported. As the studies do not require accurate evaluations of charges and particle magnitudes, certain inherent errors can be disregarded without correction, and measurements can be simplified in the interest of more data relative to accuracy of individual values.

Thus, uncertainties introduced by the irregular shapes of the particles and the consequent deviations from Stokes' Law were overlooked. The method, however, yields statistics for charge distributions for a whole cloud in relation to particle size, charge and numbers using random samples. While this method is capable of application to studies of carefully treated particles in well-controlled atmospheres, this was not attempted in the present study. Reasonable chemical and physical purification of the dusts was resorted to, and only dried room air was used. The dust clouds were dispersed by various means so as to extend from a gentle dispersion of the dusts with separation and a minimum of friction to more violent distribution with turbulence. Where needed, the blowing was controlled in as nearly a quantitative fashion as possible by means of an adjustable piston blow gun. The dispersion was also arranged so as to cover from a minimum of contact of the particles with surfaces to a maximum of such contact. Surfaces were as carefully prepared and controlled as contact with the room air permitted. Surfaces similar and different from the dusts, as well as mixtures of different dusts and pure dusts, were used. The measurements were checked as regards size distribution of the ultimate particles by microscopic examination of the dusts. This also served as a check on the absence of convection currents in the apparatus, which could seriously falsify both the statistics and charge sizes. For, if the size distribution as revealed by a microscopic study of the particles collected below the measuring field on a slide differs, especially in the region of small particles, from the distribution calculated from the rates of fall, there have been convection currents, either upward or downward, that produced the effect.

The principle of the method involved the production of the dusts by one of the various methods indicated below at the top of a settling tower of some meter in height (Fig. 1). At will, the cloud could be produced at different heights above the analyzing chamber. The cloud then settled down and along some small axial section fell through a slit into the observation or analyzing chamber containing a pair of vertical deflection plates, where it was photographed in a series of exposures on such a precalculated schedule to yield suitable sampling. The photographs indicated positive, negative and neutral particles in sizes ranging from 0.5

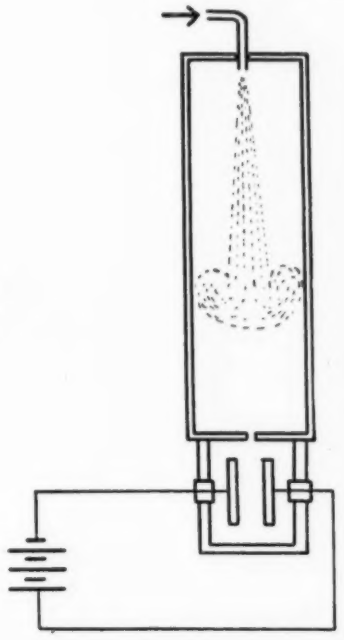


FIG. 1. Schematic drawing of apparatus showing principle of method.

[7] V. D. Hopper and T. H. Laby, Proc. Roy. Soc. **A178**, 243 (1941).
[8] J. W. Hansen, Phys. Rev. **72**, 741 (1947).
[9] W. B. Kunkel and J. W. Hansen, Rev. Sci. Inst. **21**, 308 (1950).

micron diameter to 30 microns. The charges varied from 0 to 30,000 electrons. The average charge per particle and the net charge of a sample depended on the substances used and processes of dispersion. By using large samples, these effects were reproducible. The cloud samples photographed in their measurements ranged from 200 to 2700 particles in a given cloud depending on the purpose of the investigation. The tracks of the particles observed consisted of a series of some ten dots separated by appropriate intervals, depending on the duration of the light flash and intervening dark time as set up on the rotating shutter in the optical parts (Fig. 2). Uncharged particles showed a vertical line of dots, unless the particle was largely irregular in shape. Particles out of focus by being too near the camera appear as circles; those too far away appear as stars. These appearances are predictable effects resulting from the optics. The tracks inclined to the one side or the other indicated charged particles of sign depending on the direction of the field used. The rate of vertical fall and the rate of horizontal movement allowed both charge and mass to be determined by insertion into the appropriate equations. In practice, the photographs were projected to a given scale of enlargement on a screen (usually 50×) and by means of a properly constructed nomographic scale, the charges

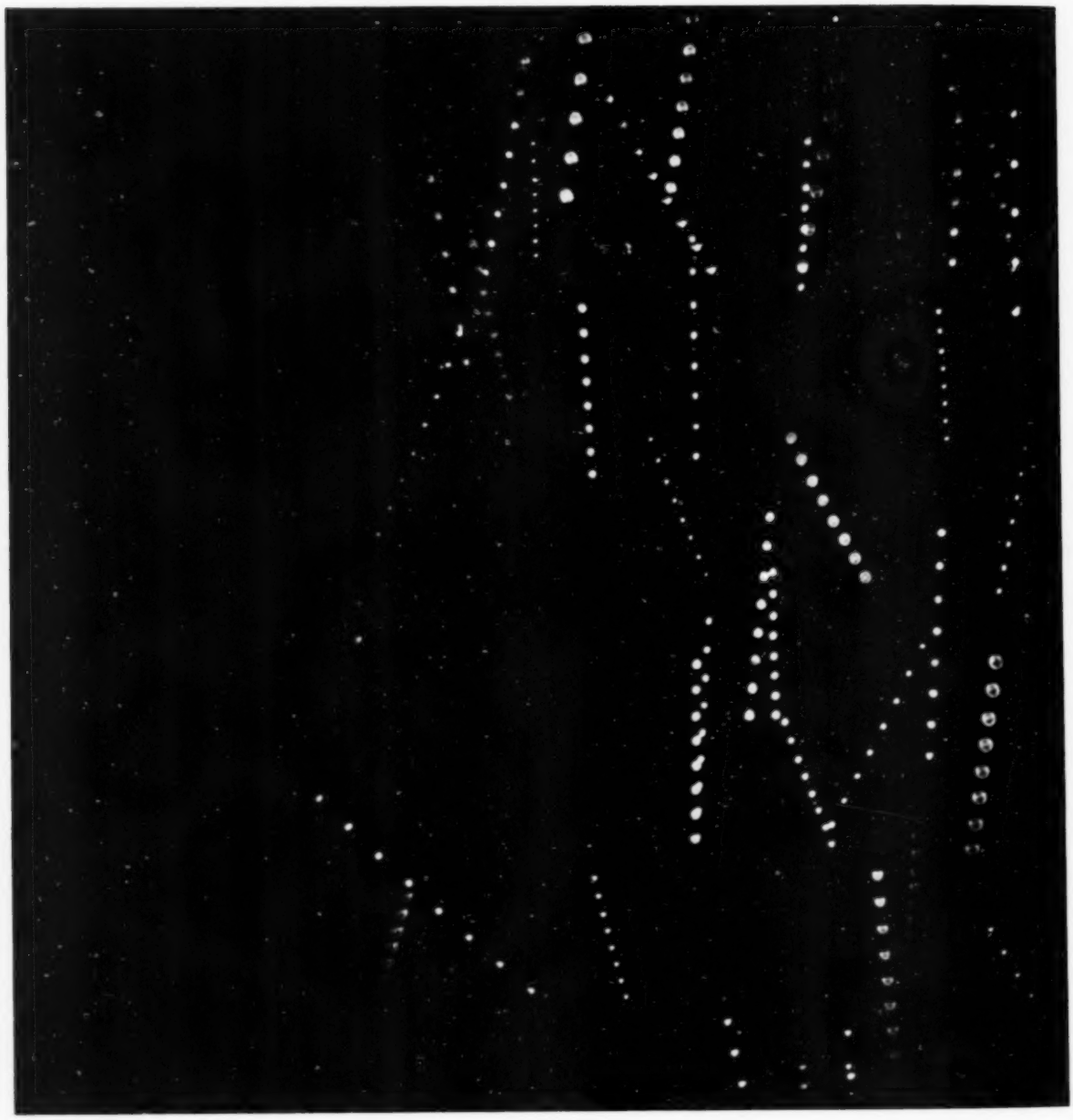


FIG. 2. Quartz particles settling in air under gravity in a horizontal electric field of 225 v cm^{-1}. 1 second exposure, 450 flashes per minute in dark field illumination. Total vertical distance shown is 5 cm.

FIG. 3. Photo micrograph of quartz powder. Most particles are seen to be very little larger than one micron in diameter.

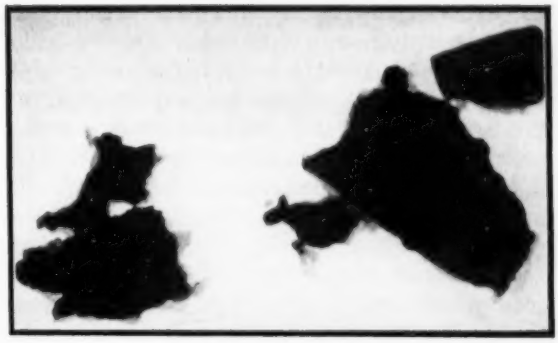


FIG. 4. Electron micrograph of quartz powder. Resolving power 50A approximately. The large particle was about four microns long.

and sizes could be inferred within a sufficient degree of accuracy for the purpose of statistics directly. When greater precision was required, more accurate data could be taken.

It must be realized that when a cloud is blown at the top of the settling chamber, there will be a segregation of particle size in falling through the chamber. Very shortly after the cloud is blown, the first very heavy particles reach the slit of the analyzer. As time goes on, the heavy particles settle out, and later—in fact even hours later—the smallest particles reach the analyzer. Thus the analysis of the cloud will consist of a properly timed and chosen set of photographs of the transient particles in the analyzer. To understand this, it is best to consider a characteristic movement.

Assuming a spatially homogeneous cloud to have been blown at a time $t=0$, the experiment consisted of a series of photographs of particle tracks taken at certain intervals. The number of particles observed in this fashion represented a certain fraction of the total number of particles in the narrow column settling through the slit at the top of the analyzing chamber. This fraction α, however, depends on the rate of settling of the particles, i.e., on the size, $\alpha=\alpha(d)$, where d gives the particle diameter in microns. In other words, the sample observed was not completely random but contained a certain size selection. Although we were not primarily interested in an exact representation of the size distribution in this study, some correction for this selectivity had occasionally to be made. A first correction could be applied by assigning proper statistical weights to the size groups observed on a picture. If W was the height of the region photographed and $v(d)$ was the rate of settling of the particles of size d, then the number of such particles included in the exposure time t is proportional to $W+vt$. The total number of such particles present in the column between successive exposures is proportional to vT where T is the average time between successive observations. Thus, the proper weighting factor for the particles of size d observed in exposures of duration t taken at intervals T is $\alpha=vT/(W+vt)$. If t and T are varied during the course of the experiment such that α for the average size observed on the pictures is kept constant, the sample taken did not falsify the actual size distribution greatly.

On the supposition that the above condition was fulfilled, the particle sizes and charges measured were tabulated and sometimes plotted. The charge-size distribution diagrams thus obtained for various samples give a good impression of the characteristics of the dust clouds in question. Frequently, the average charge per particle was calculated for successive size intervals, and the total charge present in the samples was evaluated. The results obtained and represented in this form were, in general, sufficient to provide all the desired information. In many cases, it was sufficient merely to compare the numbers of positive and negative particles in the samples and to estimate their charges only very roughly. In such a highly simplified representation of the experimental results, the element of chance introduced by the fact that only a sample of the cloud was measured could be expressed in a very simple manner. Since the particles were practically always charged and since the total clouds were always very large compared to the samples measured, the chance of finding a certain distribution in a random sample is given by the well-known binomial probability distribution law assuming homogeneity of the cloud. For large samples, such a binomial coefficient function exhibits an extremely narrow and steep maximum at the point of the actual distribution in the cloud. A half-width of such a maximum can be defined as that variation in sample distribution for which the probability is, say, one-tenth of the maximum value. If the samples are large enough, this half-width does not depend very much on the position of the maximum and is, therefore, evaluated most easily for the case of an equal distribution; i.e., for equal numbers of positive and negative particles in the cloud. This has been done for all experiments and will, henceforth, be called the standard deviation of the particular measured distribution.

The method of analysis outlined above is valid only if the cloud does not have its particles change in size or charge with time in the settling chamber. This subject

had been thoroughly investigated in all its aspects in view of some early results of Hansen.[10] The conclusions arrived at are as follows. In general, dust particles that settle quietly will stick together when they touch, thus, effectively growing less in number and larger in average diameter. A calculation has shown, however,[11] that for the cloud densities we were dealing with, the probability of such agglomeration was entirely negligible. We concluded, therefore, that the large number of aggregates invariably present in the clouds existed at least from the moment of dispersion on. This conclusion is confirmed by the fact that the amount of aggregation observed does not seem to be related to the length of the settling column.

Comparing the particle size distributions of the dust clouds for the cases of dispersion with much turbulence to those where the turbulence was slight, no significant difference could be discovered. From this result, it was concluded that the aggregates or glomerates had existed as such in the packed powders and were never separated into the ultimate particles by the stream of air, rather than that they were formed out of single particles during the violent agitation at the time of dispersion.

Another potential source of trouble is that the charges on the particles will, in general, tend to decrease in time. This arises from the presence of gaseous ions of both sign. The ions are constantly created in the air by stray radiation, and despite recombination, they will be maintained at a finite ion density in the air of the settling chamber. Highly charged dust particles suspended in the air will have an appreciable chance of catching these ionic neutralizing charges. In fact, a simple calculation reveals that for cloud densities of a thousand or more charged particles per cubic centimeter, almost all the ions created in air will attach to the dust before they have a chance to recombine. Thus, the total charge of one sign in a dust cloud will decrease linearly with time at a rate equal to the rate of ionization of the ambient air. As this rate is small, the decrease in charge is usually very slow, so that our experimental results will only be effected in the case of particles of very long settling times; i.e., for the very small particles. This is, indeed, what was observed. This problem has been investigated in detail and will form part of another paper. The results of the analysis allow of a proper computation of certain factors. Whenever an estimate of the initial charge of such extremely small particles is desired, an appropriate correction must thus be made. In addition, it should be borne in mind that the more highly charged particles will be neutralized more rapidly than those that were initially less charged. Thus, not only is the average charge of a cloud initially higher, but the initial spread of charges is larger. The corrections are not large compared to the charges involved and since little use will be made of the accurate evaluations, these corrections and their elaborate theoretical treatment will be omitted in this discussion.

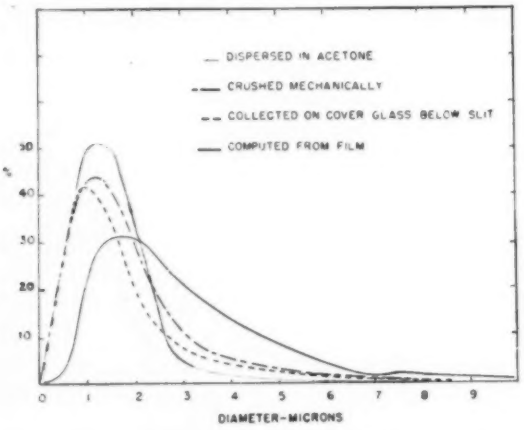


FIG. 5. Particle size distribution in precipitated sulphur. This graph shows the effect of a slight vertically downward convection in the analyzing chamber letting the particles photographed appear too large (settling faster than they should).

[10] J. W. Hansen, thesis, unpublished (1948).
[11] W. B. Kunkel, J. App. Phys. **19**, 1053 (1948).

THE POWDERS

The materials used in this study were chosen for a number of reasons. By selecting substances of widely varying physical and chemical properties, it was hoped, in the first place, to distinguish between the relatively important and unimportant variables limited, however, by three factors: the substance had to be easily obtainable in powder form of a grain size between 1 and 30 micron diameter; the powders had to be easily dispersible, with as little lumping as possible; and, lastly, the surfaces had to be reasonably clean. Unfortunately, this meant that neither hygroscopic nor chemically very active substances could be studied.

The desire for relative purity and chemical homogeneity, also, explains why no attempt was made to obtain and work with artificially rounded particles, rather than with the natural irregularly shaped substances.

The first part of this work logically undertook the study of the charges of as simple and elementary insulating substances as possible, in order to see what charges, if any, appeared on homogeneous systems. It was limited to two substances: quartz and sulphur only. These could easily be obtained in powder form of the desired grain size in fairly pure condition. The quartz was obtained by crushing and grinding. The sulphur was prepared by precipitation or crushing. The desired size range could be separated by sifting or, since neither of these are soluble in water, they could be sized by elutriation methods. Another advantage was that the containers could be made out of the same materials or in the case of sulphur, at least lined by the same material, so that sulphur only came in contact with sulphur and quartz with quartz. The main disadvantage with these substances was the particle shape, which, in general, was very far from spherical. Micrographs of

crystalline quartz are shown in Figs. 3 and 4. The electron microscope revealed that the surfaces frequently were far more irregular on a microscale than would be expected of crushed crystals. Although, in this sample most particles with diameters less than one micron had been carefully filtered out, a large number of invisible fragments were still present on the surface of the larger particle. They may have adhered to the larger particles on sizing, or they may have broken off the larger particles in the subsequent handling of the powder and then were later picked up. The presence of this coating of finer quartz dust in all observed particles must be borne in mind in the interpretation of the results. The sulphur particles were even worse in respect to regularity, with the additional drawback that they had a strong tendency to cling together very tenaciously, thus forming large aggregates of extremely irregular shapes. This could be ascribed to heavy electrostatic charging and small surface leakage. It was very difficult to break up these aggregates as shown in Fig. 5. Certainly, a gentle stream of air would be very inefficient in dispersing such a powder. This is characteristic of pure sulphur, so that in industrial practice, whenever a sulphur dust is desired, small admixtures of a suitable substance are added to prevent lumping. Since, initially, we were interested in the effect of sulphur on sulphur only, we could not make use of the dispersing admixtures. Thus, in our observations, the sulphur always contained a larger percentage of aggregate than the quartz. As indicated, the irregularity of the particles falsified the size and the charge as deduced by means of Stokes' Law,[12] and by sideways motion introduced apparent charges when none were present. To determine the magnitude of these simulated charges, test runs were made, photographing the dust as it settled under the influence of gravity alone. About 20 percent of the quartz particles showed slight deflections from the vertical, but the charges that would be calculated from these deflections in a usually fairly high field (about 200 v/cm) proved to be very small compared to charges actually appearing. Apparently, the constant tumbling motion in a field free space, due to Brownian rotation and differential heating of the settling dust particles, considerably reduced the amplitude of the helical motion,[13] which appeared to be much larger in the model study. There is evidence that in the presence of strong fields, these salutary rotations of the particles may have been absent. This came from the observation that the intensity of the light scattered by the particles fluctuated less when the electric field was present than otherwise. It is, thus, possible that, after all, the lateral drifts were somewhat larger in the actual charge measurements with the field. In the interpretation of the results, however, we shall consider these fictitious charges as negligible, inasmuch as they produce little error on multi-charged particles and, perhaps, gave a few more small lightly charged particles at the expense of neutrals than there were.

Another difficulty arose from the fact that both quartz and sulphur exist in different solid phases. It was, therefore, necessary to investigate in how far such a system could be regarded as homogeneous. The powder of precipitated sulphur was definitely rhombic, but when part of it was melted down to be used as a coating of the container, it would solidify in the amorphous phase. It was found desirable to allow at least two days to ensure that the coating had completely crystallized in the rhombic phase again. In the case of quartz, however, this procedure was not possible, since fused silica will not crystallize. We restricted ourselves, therefore, at first to crushed amorphous quartz powder in fused quartz containers, but later discovered that crystalline powder behaved in the same fashion. Either the surface of the carefully washed and thoroughly hydrated quartz powder stayed amorphous after dehydration, or the surface of the fused quartz assumed crystalline properties, or, possibly, in contrast to the sulphur, the difference between the two conditions was not sufficient to result in any detectable deviation from the homogeneous case. Consequently, the difference between amorphous and crystalline quartz was disregarded in subsequent work. Thus, the main study was done with a commercial grade of crushed quartz which was supposed to be quite pure. Since it might have been crushed in a metal mortar, it had to be washed in dilute hydrochloric acid and afterwards thoroughly rinsed in distilled water. After this treatment, the powder was dried for several hours at 120°C and stored in a closed glass jar. Nevertheless, it was found that the quartz would reach an equilibrium moisture content of a water layer of one to two molecules thickness surrounding the individual particles, which could only be driven off by prolonged heating to above 500°C. This dried powder would be so hygroscopic that it would acquire its former moisture layer very rapidly, even when placed in a dessicator containing phosphorous pentoxide. More surprising, however, was the fact that even in a very humid or almost saturated atmosphere, the water content of the powder, by actual observation, did not increase noticeably. Apparently, the average radius of curvature of these small particles was too small to permit water to condense on them in excess of the first layer that was adsorbed.

Similar observations were made on the sulphur powder, although the latter could, of course, not be baked out. An attempt to dry the sulphur at normal temperature in vacuum resulted in sublimation of sulphur, rather than in a removal of the last water layers. Only when the quartz and sulphur powders were placed in a supersaturated atmosphere could an increase in weight be detected; but in those cases, water condensed profusely on all surfaces, making the powders at once wet and lumpy. No intermediate stages of

[12] W. B. Kunkel, J. App. Phys. **19**, 1056 (1948).
[13] E. E. Miller and Tanner, Phys. Rev. **77**, 761 (1950).

moisture content seemed to be stable. Thus, the effect of water layers, if any, on the electrification of these non-hygroscopic substances could not be well-determined. It may be stated, however, that because of its relative stability, the monomolecular moisture film of such powders does not seem to be an important variable in the electrification phenomena. Certainly, the nature of these films rule out completely the old electrolytic ion exchange theory of tribo electricity.

The sulphur that was largely used in these studies was commercially produced, fairly pure precipitated powder, which happened to have just the right size distribution. Being a chemically fairly active substance, however, and moreover being produced in a chemical reaction, it was not expected to be completely free from impurities. Certainly, H_2SO_4 must have been present in spots, owing to slow oxidation in air. The insensitivity to moisture, however, indicates that these spots were few indeed.

Another substance which was chosen for its interest as a different type of powdered substance, e.g., a powdered carbohydrate, was a sample of very pure natural rice starch, the granules of which were very nearly spherical and uniform in size (between three and five microns in diameter). With this and some of the other powders mentioned, the studies had progressed so far that we no longer were interested in preventing the contact between dissimilar substances. Thus, it was possible to look at a wider range of powders which did not have to be made into coatings of the containers. Since it did not seem possible to take measurements on particles where the moisture had been completely removed and working in moisture-free systems without changing the experimental procedure, several runs were made with a talc powder, which was considered to be a naturally "dry" substance. The grain size and shape was very similar to that of the quartz powder.

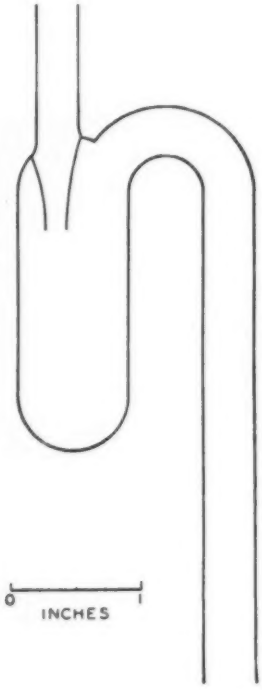


FIG. 6. Dusting tube made of quartz or Pyrex, sometimes lined with sulphur.

Obviously, among the simple elementary dusts, metallic powders come to mind. These are, however, very hard to produce in a satisfactory condition. In order to study the behavior of metallic particles, a nickel powder was prepared by rubbing a nickel rod with emery paper and separating the metal particles from the abrasive by means of a magnet. The powder had to be demagnetized and reduced in a stream of hydrogen. After this treatment, it appeared to be readily dispersible and of sufficiently fine grain size. The shape of these particles was, of course, extremely irregular, especially the larger ones, but since we had a nickel dust disperser, chosen because of its relative resistance to corrosion, it was considered desirable to investigate the behaviors of a powder of the same material. (It was only after the experiments had been discontinued that we were able to obtain a sample of the nearly spherical carbonyl iron powder. Such a powder, however, would be more heavily oxidized and, therefore, probably not representative of a metal surface.)

METHODS OF DISPERSION

In the interests of simplicity of operation, it was decided to employ dispersion by short blasts of air only. Since it was essential to the measurements that violent air disturbance in the settling tower be avoided, only a relatively small amount of air could be applied. For the same reason, more vigorous methods of cloud generation like explosions or mechanical shaking had to be discarded. We, thus, were rather limited in the range over which the factors involved could be varied. Nevertheless, as will be shown, this small range proved to be sufficient to permit the most essential conclusions.

In order to distinguish between some of the various mechanisms by which the charge separations might be effected, it seemed plausible, as a first step, to compare the result of a maximum of friction with that of negligible friction.

Hence, the simple dusting tube shown in Fig. 6 was designed so that even a gentle stream of air would cause strong turbulence, thus offering a maximum chance of frictional contact among the whipped up particles themselves and between the dust and the walls. Such tubes could easily be made of fused quartz or of Pyrex. One of the Pyrex tubes was carefully lined with sulphur. A similar duster was assembled out of nickel tubing. On the other extreme, it was impossible to disperse any of the powder without causing some slight friction between particles. However, it was attempted to reduce the contacts between particles in such dispersions after the initial separation to a minimum by blowing the powder very gently out of a small thimble-like container as indicated in Fig. 7. The motion in the clouds

thus produced indicated that turbulence was negligible in this case. The cups again could easily be made out of quartz or sulphur to avoid any possibility of contact between dissimilar substances when it was not desired. When, on the other hand, the effect of collisions of the dust particles with metal walls was to be studied, a third type of dusting tube was constructed especially designed to offer a large opportunity of particles striking the sides, Fig. 7. The number of baffles could be varied up to a maximum of fifteen per tube. The idea of forcing the stream of air into a zigzag path with sharp bends is fairly obvious and seemed to work well in many cases. The sample to be dispersed was placed on the second baffle from the top. Two of such tubes were constructed, one of nickel and one of platinum, avoiding all traces of foreign metals on the inside.

Early in the work, it was realized that the amount and velocity of the dispersing air stream had to be rather accurately controlled if reproducible results were expected. While the amount of charge appeared to be insensitive to the air blast, the amount of dust and thus the cloud density varied widely with the air stream used. Since these had to stay within certain limits to facilitate analysis and make statistical interpretation possible, it was clear that the air blasts had to be quantitatively reproducible. To achieve this, a blowing cylinder was devised, consisting of a brass tube of 6.25 cm diameter with a plunger bearing rubber "O" rings. This air gun had to be mounted upright above the dusting tube or cup. The outlet was at the bottom carrying a stopcock and a long thin detachable nozzle. The plunger could be lifted to entrap the desired amount of gas in the cylinder, the pressure could be regulated by loading the plunger with the proper weights. A further advantage of this design was that the blast to be applied did not necessarily have to consist of ordinary room air, but could consist of dried air or any gas which did not react with the tubing. This device performed very satisfactorily.

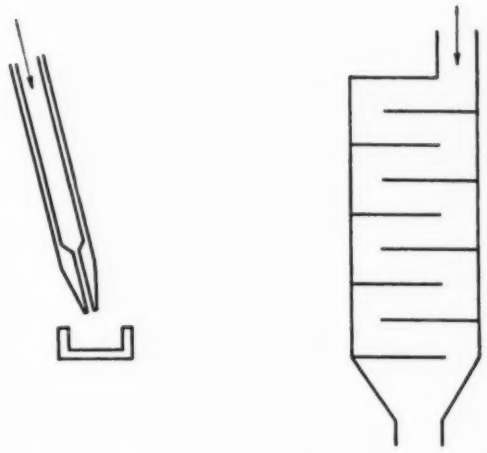


FIG. 7. Left: small cup and nozzle, made of quartz or sulphur. Right: metal duster made of nickel or platinum. The sample of powder to be dispersed was placed on the second baffle from the top.

THE RESULTS

All in all, more than eighty separate experiments were performed in this study, consisting together of more than 3000 exposures. Some 50,000 particles were photographed. The charge and size of not all of the particles were measured accurately. A large fraction of the charges and sizes were just estimated very roughly. The frequent repetitions of experiments under identical conditions showed the results to be completely reproducible. There is, therefore, little doubt that the findings here reported are reliable and of a general significance.

It must first be noted that every single cloud investigated *contained both positive and negative particles simultaneously*. Whenever any contact between unlike substances was completely avoided; e.g., blowing quartz powder out of a quartz tube or sulphur dust out of a sulphur cup, the numbers of positive and negative particles were about equal. Dividing the particles into groups of certain size ranges and calculating the average charge per particle for the various sizes, it was found that these were also equal for the two polarities. The average charge per particle increased, in general, somewhat more slowly than the square of the diameter, i.e., not quite linearly with the surface. This observation was quite general in all studies. Very few of the particles observed, in the range of sizes studied, were uncharged.

A good description of the charge, size and relative frequency distribution of the particles in such a dust cloud can be given graphically by representing each particle observed by a dot at the appropriate point in a charge-size plane (Fig. 8 for blown quartz dust). Since, in the particular experiment shown, more than 1500 particles were measured, only every fifth one was plotted in order not to overcrowd the graph. The averages drawn in, however, were computed from the total number. The picture indicated in Fig. 8 was typical of all of these homogeneous cases. Even the average charges never differed very much in these homogeneous studies, despite much variation in dispersion procedure. This is a rather remarkable result, since the amount of friction or the number of impacts between the particles presumably varied appreciably, depending on the amount of turbulence during dispersion. Particular values for fine quartz dust blown out of a quartz duster and out of a small cup, representing the extremes studied, are given in Table I. Similar values were found for sulphur dust. Several explanations of this unexpected behavior suggested themselves. The processes by which charge separation occurred apparently were not influenced by the difference in turbulence. This probably means that the charge separation occurred at the first separation of the contiguous particles and the charges only possibly redistributed themselves slightly at subsequent collisions which would not affect their magnitude appreciably. Whether these charges were

TABLE I.

SiO_2 Size Micron diameter	SiO_2 duster Number of particles +	0[a]	−	Average charge +	−	SiO_2 cup Number of particles +	0[a]	−	Average charge +	−
0–1	147	4	161	15.3	18.2	158	0	157	13.9	14.1
1–2	446	32	474	22.2	23.2	340	8	409	23.4	25.2
2–3	84	5	96	45.8	49.3	127	6	125	52.8	59.8
3–4	29	3	30	72.4	67.3	36	1	35	103.8	86.8
4–5	7	0	8	112	99.2	21	3	23	134	172
5–6	5	0	4	105	117	15	1	9	155	159
6–7	0	0	2		342	7	2	6	141	381
	Total numbers			Total charges		Total numbers			Total charges	
	718	44	775	20,000	23,000	704	21	764	27,000	31,300

[a] Here particles were considered neutral when their charges were too small to be determined; i.e. $q<3$ for $d\approx 1\mu$ increasing to $q<30$ for $d\approx 10\mu$.

already irregularly distributed in the packed powder, or whether the separation occurred at the moment of dispersion could, of course, never be detected by our method. But, since the powder as a whole was neutral, and since any charge is effective only when separated from its opposite, this question is of little interest. We could, also, assume that the charge distribution reached some kind of an equilibrium after comparatively little friction, so that any further impacts and additional friction would not change the resultant charge distribution noticeably. The difference between these two interpretations can be summarized in the following way: In the first cases, the main charge separation is supposed to occur at the first mechanical separation of the particles; whereas in the latter, it is regarded as a result of several collisions and possibly friction of very short duration. However, as later interpretation shows, the evidence is all in favor of a charge distribution on separation of contacting surfaces. There is, of course, a small chance that the strong variation in turbulence, contrary to expectation, did not effectively change the amount of friction and the number and violence of the impacts between particles. In that case, it would be expected that no difference in the observed particle charges would be found. At any rate, the electrostatic cloud characteristic shown in Fig. 8 was surprisingly independent of the conditions of dispersion.

Changing the moisture content of the surrounding atmosphere in the hope of varying the water layers on the dust particles, and thus the charging, produced no material change in the type of distribution shown in Fig. 8. But, as already pointed out, the water content of a fine non-hygroscopic powder was known to be remarkably unaffected by the relative humidity of the surrounding medium, so that any significant variation in charging due to moisture was hardly to be expected.

A short summary of the most important experimental results is given in Table II.

There it is seen that still more surprising than the invariance of the charges generated in homogeneous systems, was the fact that many so-called inhomogeneous systems showed nearly the same symmetric charge as in Fig. 8. When quartz, talc, or starch was blown out of a pyrex duster, the charges on the particles balanced just as before* and when sulphur was blown out of a quartz, Pyrex or nickel tube, the predominance of one sign of charge was only very slight.

Again, two different explanations are possible. First of all, there was little doubt that only a small fraction of particles actually hit the solid walls in the process of generation, most of them being surrounded by a layer of air all the time, and thus being cushioned by the air and failing to strike the walls. Larger particles had a better chance of impact than smaller particles because of their inertia. Thus, most of the charges were generated by dispersal into the cloud or by collisions between particles. Those particles that hit the walls either did not interact sufficiently with the wall lattice to cause any appreciable charge transfer, or else the differences in surface characteristics were not large enough on the average to disturb the initial symmetry perceptibly. Strangely, the sulphur powder did not show a very large asymmetry, even when forced through the nickel duster with as many as eight baffles. In that case, certainly, most of the larger particles must have hit the sides. We have not been able to account adequately for this fact, except, possibly, by noting that the observed samples could not be made large enough to yield reliable statistics. (The sulphur was not easily dispersible.)

The talc powder, when blown through the nickel chamber, did not show the slightest asymmetry in its

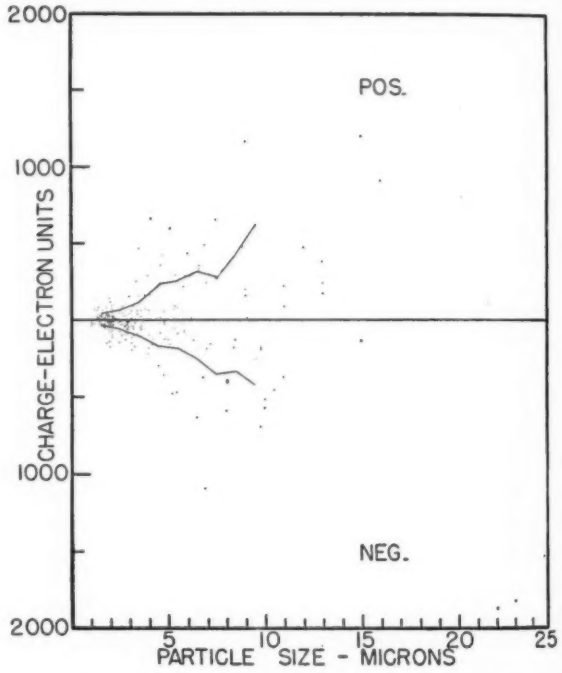


FIG. 8. Charge *vs.* size representation of a cloud of quartz powder blown out of a quartz cup. The average charges per particle are drawn in; the spread is seen to be large.

* A similar result was obtained by Hansen with $CaCO_3$ powder in a SiO_2 duster.

charge distribution. Here, however, the explanation was obvious. Opening the nickel duster after the experiment revealed that the inside was completely lined with a heavy layer of talc powder. This layer could not be blown off but had to be wiped away. Apparently, all talc particles hitting the metal sides stayed there, possibly attracted by their image charges, until all metal surfaces were covered, thus effectively reducing the conditions to those of the homogeneous cases.

The only fairly pronounced asymmetry in charge distribution involving solely non-conductors was obtained when a fine quartz powder was blown out of a sulphur lined duster. The negative sign predominated distinctly, which was in qualitative agreement with the fact that in the sulphur powder dispersed out of a quartz duster the positive charge was noticeably larger.

It was noted, at the same time, that the asymmetry was stronger in the larger particles, being very slight for the smaller sizes. All other strongly asymmetric cases contained metallic elements. When insulating powders were used with metal or vice versa, this insulating dust invariably acquired a predominantly negative charge, leaving the metal with the corresponding positive balance. The largest deviation from the symmetrical distribution occurred in a quartz cloud blown out of the platinum dispersing chamber with eight baffles. A plot showing every other particle is given in Fig. 9. It is seen that the negative charges predominated both in number and magnitude. Still a large number of positive particles, especially of small diameter, were present. None of the larger particles carried a positive charge. Evidently the quartz in contact with platinum was likely to acquire a surplus of electrons.

All large particles or glomerates hit the metal surfaces at least once, whereas many of the smaller ones never touched the walls, being carried along by the air stream well imbedded in the gas. Many of the small particles that succeeded in colliding with the sides of the chamber stuck there, forming a thin coating. A crude measurement of the positive charge imparted to the duster showed this to be of the same order of magnitude as the total net charge estimated to be present in the cloud. This can be considered as evidence that the observed negative surplus charge was not balanced by invisible positive particles left behind in the settling column. If desired, this conclusion could, of course, very easily be verified by employing any of the many well-known space charge detecting devices.

In agreement with these findings, a nickel powder blown out of the quartz dusting tube carried a predominantly positive charge with a spread otherwise similar to that shown in Fig. 9. Either the nickel particles, unlike those of sulphur, had a large chance of touching the sides of the tube, possibly because of their larger inertia for equal volumes (higher density), or the charge transfer between nickel and quartz is easier than that between sulphur and nickel. The nickel powder, when dispersed out of the nickel duster, however, did not behave like the other homogeneous systems, very probably because the surfaces of the reduced Ni particles differed appreciably in their properties and constitution from the nickel plate walls. This was despite the fact that every common precaution possible had been taken to insure uniformity and zero contact potential.

TABLE II.

Symmetric					Asymmetric				
Powder	Container	Sign	Deviation[c]	St. dev.[c]	Powder	Container	Sign	Deviation[c]	St. dev.[c]
SiO_2	SiO_2	−	0.035	0.05	SiO_2	S	−	0.21	0.11
S	S	−	0.036	0.04					
Starch	Pyrex	−	0.020	0.12	Ni	SiO_2	+	0.21	0.07
Talc	Pyrex	+	0.015	0.08	Sb[b]	Pt	−	0.60	0.21
SiO_2	Pyrex	−	0.10	0.12	Ni	Pt	+	0.34	0.10
Talc	Ni[a]	−	0.06	0.11	Ni	Ni	+	0.32	0.09
					Starch	Ni	−	0.33	0.11
Slightly asymmetric					SiO_2	Ni	−	0.39	0.07
S	SiO_2	+	0.13	0.11	SiO_2	Pt	−	0.50	0.11
S	Pyrex	−	0.13	0.11					
Sb	Ni	−	0.19	0.16					

[a] Ni duster found coated with a layer of talc powder.
[b] Sulphur powder could not be well dispersed.
[c] The deviations from symmetry are expressed by the ratio $x=\frac{N_+-N_-}{N_++N_-}$.

As shown in Table II, an observed distribution is considered to be symmetric when the deviation from the exactly balanced distribution is less than the statistical standard deviation of the particular experiment. Similarly, if the deviation is only a little larger than the standard deviation, the distribution is called slightly asymmetric. If it is much larger, the result is regarded as asymmetric. In this way, the statistical factor, which is unavoidably connected with any measurement on random samples, is at least partially taken into account. This procedure of classifying the results merely according to the relative numbers of positive and negative particles instead of according to the magnitude of these charges is justified, since all experiments show a strong correlation between the relative numbers and the average charges of the particles of opposite sign.

Finally, in order to demonstrate most convincingly that much of the charge separation takes place during the very first phase of the dispersion, the fine quartz powder was mixed with the easily lumping sulphur and blown very, very gently out of a quartz cup by a slow vertical stream of air. In this case, the large particles seemed to be fairly randomly charged with only a slight excess of the negative sign; whereas the small particles which arrived later proved to be overwhelmingly negative. These latter were certainly mostly quartz; the former must have consisted partly of the large sulphur glomerates. Since both the contact between particles and dissimilar walls and the turbulence and friction between particles was presumably negligibly small, the charge separation must have become effective at the very first breaking of the close contacts between the dissimilar surfaces of the quartz and sulphur particles. It should be noted that in

cases like these, all elements of a static machine separating or segregating charges which could ultimately lead to sparks are present. The dissimilarity of large and small particles results in an asymmetry of charge by small particles having one sign larger than the other. Mechanical separation in the gravitational field or by air currents then leads to segregation of charges and local accumulation.

CONCLUSIONS

There appears to be no doubt that the act of separating a mass of particles of solid in contact with each other into a cloud of assorted sizes of particles, yields in all cases a cloud which as a whole is neutral, but in which practically all the particles are charged. In a homogeneous substance, where contact with heterogeneous surfaces are at a minimum, the numbers of positive and negative particles in each size group from about 0.5 micron to 30 microns in diameter are about equal. The average charge increases somewhat more slowly than proportional to the surface of the particles. Apparently the character and forces involved in the separation and the possibility of subsequent friction in turbulent clouds with violent blowing produce no notable change in the electrification, nor does the charge increase with increase in contact or the energy involved in the process. In the separation of an originally mixed dust of different average size for the two components, the study of the cloud indicates that the electrification occurs primarily at the instant of separation of the two components. If homogeneous clouds are blown so that they make contact with solid surfaces of a different sort in sufficient measure to affect a notable fraction of the particles, the separation of the heterogeneous system of dust and solid surface may impose an asymmetry on the equality of positive and negative particles from the dispersed cloud particles that did not contact the solid surface. Results are reproducible in any given system. However, in a surprisingly large number of impacts between dissimilar surface and dust, no great influence was observed. In general, the asymmetry is exhibited more strongly by the larger particles, which, because of their mass, were more likely to strike the solid surface. Very small particles of the order of one or two microns appear to be cushioned by the gas and do not make contact with the surface in some cases. In occasional cases, e.g., talc on metal, care must be taken in interpreting the results, since the dust may be so highly charged and insulating that it is held to metal surface and most of the cloud makes contact with the dust coating only, showing the charge distribution of a homogeneous substance. The direction of charge transfer between heterogeneous surfaces is unpredictable and depends on the substances involved. With some insulators, the character of the surface, whether crystalline or amorphous, appears to cause no asymmetry or charging. Thus, amorphous quartz and crystalline quartz showed no asymmetry. Likewise, Pyrex glass, a borosilicate glass, and quartz showed no asymmetric transfer. The most universal asymmetry observed was that of metal and insulators, where the insulator universally received a negative charge. In the case of a Ni powder reduced in H_2 against Ni or Pt sheet strong asymmetry was observed with the reduced Ni powder losing electrons to Pt and to the sheet Ni. This is to be expected in the case of Ni and Pt, for Ni has a low work function relative to Pt. In some cases where asymmetry, owing to contact of dust and surface, predominated, the charge on the duster roughly equalled the total net charge of the opposite sign on the cloud. These results, as a whole, and especially those with metals, not only through their consistency and reproducibility, but through the signs of electrification obtained, give considerable confidence in the use of this method and the results it yields. In fact, it now appears that the method is ready to be applied to systems of exceptional purity and in carefully controlled atmospheres, in order to derive some hints as to the true character of the processes of charge exchange in non-homogeneous systems as related to the solid state.

Insofar as these studies go, they were done on reasonably clean surfaces and a fairly representative spread of substances to cover the field. The ambient gas was fairly dry and filtered air, subjected to the natural ionization by radioactivity in the surroundings and cosmic rays in the settling tower. Probably, in this connection, one of the most striking discoveries in the

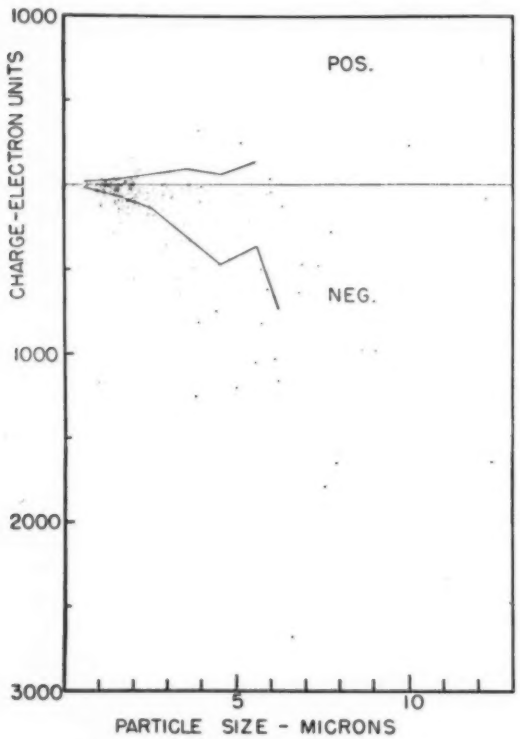


FIG. 9. Charge *vs.* size representation of a cloud of quartz powder dispersed out of a platinum duster. The average charges per particle are drawn in; note the large spread.

work with the normally dry dusts used, was the minor role played by the water vapor and moisture. Since trouble from this source was expected, considerable time was devoted to its study. Actual studies, especially on the quartz, indicated that if heated above 500°C, it lost all superficial moisture: On cooling to room temperature, or even to 120°C, it rapidly absorbed a small amount of water. By weight, it was equivalent to no more than enough to cover the average particle if spherical with, at most, a layer two molecules deep. From the electron microscope pictures of the particles of quartz, the surface is really so much greater than assumed that the layer could not be more than one molecule thick, and then the H_2O molecules may only occupy preferential spots. Changing relative humidity up to 90 percent at 20°C did not alter the layers at all. Above this, the surface became wet and the dust stuck together. Such dust did not disperse. From this it must be concluded that the water played a minor role in the electrification. Certainly, there was not enough water to give aqueous films, so that the classically conceived electrolytic ion exchange could account for the electrification.

It is probably too early to try to evolve a picture of the processes at work. However, the universal symmetry of charging on separation of homogeneous particles after contact and the asymmetrical charging of some of the heterogeneous systems, together with observations in the electron microscope, permit us to formulate a tentative theory of what happens. In the case of the homogeneous system, it is clear that the major portion of the charging comes from the separation of particles initially in contact. The very great sub-microscopic roughness of the surfaces which was observed, together with no evidence that this roughness constitutes a fine sub-microscopic dust which charges independently of the microscopic particles in the cloud, leads one to expect relatively strong pressures in contact points between the major aggregates microscopically observed. These pressures could *eliminate gaseous or surface films* and thus constitute intimate contact *between small areas of the homogeneous substance*. The separations on dispersal may take place along the actual region of contact, or they may take place by the actual fracturing of the small contacting points, depending on how intimately the pressure has welded the contacting points of the two larger pieces together. In any case, separation and dispersal into a cloud involves producing some sort of rupture between a large array of minute points. Doubtless, these points are of various sizes. If now one makes the assumption that for a contact area of some minimum size, i.e., radius σ involving, say, some 10^3 or 10^4 atoms of the crystal lattice, there is a good chance that crystal imperfections will leave one extra electron or ion of negative charge lying near the plane of fracture on one side of that plane; then that particular point of contact on breakage will contribute a negative charge to its large particle and the equivalent positive charge to the other particle. Such chances of separation are easily possible, due to imperfections in the lattice in which a negative ion or electron is out of place and is geometrically separated from the equivalent positive charge relative to a cleavage plane. Cleavage along such a plane will then cause a charge separation.

On this basis, statistical analysis enables us to compute the probability that a single large particle will have a certain excess charge, positive or negative. The larger the surfaces of the large particles in contact, the more points are ruptured and from pure probability considerations using the concept of a minimum area per unit charge imperfection, the statistics and relations may be derived.

If the powder contains a continuous spectrum of sizes, the contact area of a certain particle with its neighbors will be on the average proportional to its surface.

Let us define a *unit contact* as a region of extremely high probability (certainty) of transfer of one electron on particle separation. Thus, at each unit contact one particle is left with an excess and the other with a deficit of one electron of charge. Let us further assume that the number of such unit contacts is large on all particles to be considered and that the direction of charge transfer is completely random in any homogeneous powder.

Let the number of unit contacts of a certain particle be N. Thus, $\bar{N} \propto S$ (S=surface area of particle). If the number of electrons gained is K ($K<N$), then the number of electrons lost at the other contacts is $N-K$ and the net charge finally will be $n=(N-K)-K=N-2K$ elementary charges. The probability of such an event is given by

$$P_K{}^N=\frac{N!}{2^N(N-K)!K!}=\frac{N!}{2^N}\cdot\frac{1}{(N/2+n/2)!(N/2-n/2)!}.$$

(This assumes exactly equal chances of electron loss and electron gain.) Since N is large, we can use Stirling's formula and obtain

$$P_n{}^N\cong\left(\frac{2}{\pi N}\right)^{\frac{1}{2}}\frac{1}{[1+(n/N)]^{(N+n+1)/2}[1-(n/N)]^{(N-n+1)/2}}.$$

If N is large, such a binomial distribution curve can be very adequately approximated by $P(n)=(2/\pi N)^{\frac{1}{2}}\times\exp(-n^2/2N)$. The plot shown in Fig. 8, indeed, suggests such a Gaussian distribution of charge for constant particle size. The average charge will then depend in a simple fashion on the size as can be seen from the following discussion.

$$\bar{n}=(2/\pi N)^{\frac{1}{2}}\int_0^\infty n\exp(-n^2/2N)dn=(2N/\pi)^{\frac{1}{2}}.$$

Let $N=(\pi/2)\alpha^2S$. Then $\bar{n}=\alpha(S)^{\frac{1}{2}}$. Here α is a constant factor, assuming that the statistical relation between N and S does not disturb the average. Thus, this theory

leads to the conclusion that the average charge per particle of one sign should be proportional to the square root of the particle surface.

For spherical particles $\bar{n}=Bd$ with d the diameter. Since, for the cases investigated, the particle shapes were very far from spherical, no such simple relation can be expected. Inasmuch as the irregularities always tend to increase with size, owing to the formation of aggregates, we should expect $\bar{n}$ to vary somewhat more rapidly than linearly with d which is exactly what is observed. This simple theory of random charge transfer at points of close contact then leads to conclusions in good agreement with observations. It must be noted that while the Gaussian distribution for the symmetrical cloud as deduced has a maximum at zero charge, the observations on, say, 100 particles in a given size range with some 1000 charges would not show the large number of uncharged particles of zero charge on purely satistical grounds. If this picture is accepted, the values listed in Table I will at once yield an estimate of the number of unit contacts on particles of a certain size. Thus, for the particles of diameters between 2 and 3 micron, the data give N as approximately $N=\pi/2\bar{n}^2 \cong 4000$. For those particles between 0.5 and 1 micron, the value of N is approximately $N\cong 400$.* Since, on the other hand, the surface area of the particles is known, at least very roughly, the maximum average size of a unit contact, as defined above, can be estimated. It is found from our data that for the quartz used by us, at least one electron is transferred at each area of contact of 10^{-10} cm^2. Similar magnitudes occur for all the particles observed. The differences existing are hidden in the uncertain knowledge of the surface areas involved. These values are maximum estimates. Actually, the average inhomogeneity at surfaces can well be confined to smaller areas, since only a small fraction of the total surface will be in actual contact. If the contacts are confined to 1/100 of the total surface, the average area of a unit contact will be 10^{-12} cm^2, which corresponds to about one imperfection for every 10^4 atoms, a figure that is more likely than one in 10^6.

In the case of heterogeneous surfaces, differences in structure, as well as in surface layers, will result, on the average, in charge transfers at contact points which can well be asymmetrical. This would be particularly likely where metals having free electrons can give these off to the ionic and other lattices having electro-negative elements. Even then, many of the contacts can be separated by fracture in the homogeneous substance leading to more homogeneous charging. Therefore, it will be expected that even in the most heterogeneous systems, the asymmetrical charging will never be complete. This is especially true, of course, where friction with dissimilar surfaces, as in the duster with many baffles, does not occur. It is probable, in view of what has been said, that it will be found that in heterogeneous systems the gas films and surface conditions play an important role, whereas they do not do so in the more homogeneous systems where possible pressure fusion and rupture of the minute points is more likely.

I am indebted to Professor L. B. Loeb, at whose suggestion this problem was undertaken and under whose supervision it was carried out, for his many valuable suggestions and assistance in interpretation, which were a constant source of inspiration. Also, I wish to thank Mr. R. J. Wijsman for his stimulation and verification of the statistical theory of mutual charging here presented.

* The observed average charge per particle was not corrected for neutralization by air ions in this case. Small particles, which were often photographed at a considerable time after their separation from the powder, would be expected to have lost some of their initial charge. Thus, the true value of N can be expected to be somewhat larger.

Charge Distribution in Coarse Aerosols as a Function of Time*

W. B. KUNKEL
Department of Physics, University of California, Berkeley, California
(Received April 3, 1950)

The interaction of small particles suspended in air with the ions normally produced in the air is discussed. The neutralization of highly charged dust particles, if suspended in air, is treated in detail. The result is found to be in qualitative agreement with experiment. The special case of final equilibrium is investigated. It is found that multiply charged particles should be present in appreciable number in coarse aerosols at all times. This is confirmed by experiments both for initially charged quartz dust and for initially neutral ammonium chloride smoke. The general nature of an equilibrium charge distribution is presented. Charges up to ten electron units are found to be not uncommon.

INTRODUCTION

IT has long been known that microscopic and submicroscopic particles and droplets suspended in air frequently carry electric charges.[1] Among others, Patterson in his extensive studies of aerosols[2] noted the fact and developed a theory involving the coagulation taking place in fresh aerosols of high density. Experimentally, Patterson and his collaborators found that the percentage of charged particles in originally uncharged clouds increased with time, the numbers of positives and negatives remaining equal throughout. Originally charged sols would remain charged. The magnitude of the charges were, however, not determined.

The processes by which these charges are acquired and/or altered can be divided into three groups. First, the violent creation of an aerosol, be it by a chemical reaction, by an electric arc, by the mechanical action of an explosion, or merely the dispersion through a blast of air, will in general leave the particles charged.[3] Such charging may be by triboelectric actions, by spray electrification, or possibly by direct ionization or thermionic emission due to strong local heating. These actions will determine the initial charges of the particles. Secondly, there may then be subsequent collisions between particles involving frictional charging, charge transfer and agglomeration. Such contacts caused by Brownian motion and turbulence in the medium can appreciably alter the initial distribution of charges in time. Finally, there will always be direct exchanges with the ions present in the ambient air, thus effecting a change of charge on the particles even when there is no contact between particles.

This study concerns itself with the effect of ionic charge exchanges of relatively large particles which are practically unaffected by Brownian motion, and it is limited to cloud densities small enough to render the probability of mutual impact of the particles negligible.

CHANGE OF CHARGE OF SUSPENDED PARTICLES

A small particle suspended in air is continually bombarded by gas molecules. Among the molecules there are a certain number of gaseous ions of both signs produced continually by various processes. These ions will have a finite chance of attaching to the particle or at least of being neutralized on impact, thus transferring their charge to the particle. If the particle was initially charged, this change of charge by one unit may represent either an increase or a decrease. The probability of such an event will depend primarily on the initial charge of the particle, on the ion density in the air, on the size, shape and material of the particle, on the temperature, and, if present, on any relative directed motion of the ions; e.g., on the average drift velocity of the ions, if there is an electrical field acting as in Millikan's oil drop study.

It is also possible for a particle to change its charge due to an ionizing process directly affecting the particle itself; e.g., caused by some radiation striking it, and ejecting one or more electrons. The probability of such an event will depend on the size and material of the particle, the total surface of the particles relative to the absorbing cross section of the gas molecules, and on the character, energy and amount of ionizing radiation. Owing to the values of the relative numbers of molecules to those of the particles, the direct ionizing effects for many types of radiation will be much smaller than the charging because of the ionization of the air. It is clear that the problem presented in charging of dusts by natural ions in the air as indicated above involves an equilibrium between the ions of both signs present in relatively small equal amounts, a problem apparently not solved before. On the other hand, the accumulation of charge on particles suspended in regions of relatively high ion densities of one sign as affecting electrical precipitation has been studied by many authors, notably by Arendt and Kallmann.[4]

In order to study the capture of ions in the air by particles such as are under discussion, we proceed as follows. Let us call $\varphi_1(q)$ the probability that a particle of charge q and of a certain size, shape and substance

* This work was supported by an ONR Contract.
[1] H. Freundlich, "Kapillarchemie," Akad. Verl. Ges. (1922).
[2] Patterson, Proc. Roy. Soc. **A124**, 523 (1929). Also Phil. Mag. **12**, 1175 (1931).
[3] H. Sachsse, Ann. d. Physik **14**, 396 (1932).
[4] Arendt and Kallmann, Zeits. f. Phys. **35**, 421 (1925).

suspended in a medium of a certain ion density and temperature gains an electron unit of charge in unit time. And let $\varphi_2(q)$ be the probability that an identical particle under identical conditions loses an electron unit of charge per unit time. Furthermore, let N_q be the number of such particles suspended per cubic centimeter. Then the change in the density of such particles in still air due to charge transfers, neglecting the settling motion, becomes:

$$\Delta N_q=[N_{q-1}\varphi_1(q-1)-N_q\varphi_1(q) -N_q\varphi_2(q)+N_{q+1}\varphi_2(q+1)]\Delta t. \quad (1)$$

If particles of various sizes are present in the volume element similar equations hold for each infinitesimal size range separately. Since agglomeration and evaporation or condensation are considered to be negligible, there is no change of particle size and these equations are independent of each other. Yet, inasmuch as φ_1 and φ_2 are generally unknown, Eq. (1) is of no great importance. Only in very special cases will certain simplifications enable us to make predictions.

NEUTRALIZATION OF A CHARGED SMOKE

If the individual particles of a cloud of smoke or fog are highly charged, and if the densities of positive and negative ions are of the same order of magnitude, φ_1 becomes negligible compared to φ_2 which will consist almost exclusively of the probability of neutralization by air ions. Moreover, since $q\gg1$, we can treat the changes of charge as infinitesimals. Thus we can write

$$\frac{\Delta N_q}{\Delta t}\rightarrow\frac{\partial N_q}{\partial t}dq=-N_q\varphi_2+\left(N_q+\frac{\partial N_q}{\partial q}dq\right)\left(\varphi_2+\frac{\partial\varphi_2}{\partial q}dq\right)$$

or

$$\frac{\partial N_q}{\partial t}=N_q\frac{\partial\varphi_2}{\partial q}+\varphi_2\frac{\partial N_q}{\partial q}=\frac{\partial N_q\varphi_2}{\partial q}, \quad (2)$$

where $\varphi_2(q)=-dq/dt$ is the average rate of loss of charge of a particle of charge q.

Equation (2) is of practical value only if $\varphi_2(q)$ is a relatively simple expression. To obtain a first approximation, we can assume that, on the average, ions, due to their heat motion, entering a sphere of radius r around a particle of charge q will be trapped, r being given by the relation

$$E_{\text{pot}}+E_{\text{kin}}=\frac{qq'e^2}{r}+\frac{3}{2}kT=0. \quad (3)$$

Here e is the electron charge, q the number of electron charges on the particle, q' the multiplicity of the ion charge, k the Boltzmann constant, and T the absolute temperature. The quantity r is positive only if q and q' are of opposite sign. This relation assumes a Maxwellian velocity distribution of the ions. Then the number of neutralizing events per second will be given by the well-known relation

$$\varphi_2=-dq/dt=n\pi r^2\bar{c}=g(t)q^2 \quad (4)$$

again to a first approximation. Here n is the density, $\bar{c}$ the average velocity of the ions, and $g(t)$ is a factor independent of q derived from Eqs. (3) and (4)

$$g(t)=\frac{4\pi\bar{c}e^4q'^2}{9k^2T^2}n(t)=\beta n(t). \quad (5)$$

It should be noted that in this case φ_2 is independent of the size of the particle as long as its radius $\rho<r$, so that the treatment will hold even if a cloud contains particles of various sizes as long as the cloud density does not change due to the difference in settling rates.

Substituting (4) and (5) into (2) we obtain

$$\frac{\partial N_q}{\partial t}=\beta n\frac{\partial}{\partial q}(q^2N_q). \quad (6)$$

This means that q^2N_q is independent of time if $q(t)$ is regarded as the charge on a particular particle changing according to (4):

$$q=\frac{q_0}{1+\beta q_0\int_0^t ndt}. \quad (7)$$

Thus

$$N_q(q,t)=(q_0^2/q^2)N_q(q_0,0), \quad (8)$$

where $N_q(q,0)$ is the initial charge distribution in the cloud and q_0 the initial charge of a particle of charge q at time t:

$$q_0=\frac{q}{1-\beta q\int_0^t ndt}. \quad (9)$$

The problem is therefore reduced to that of finding $n(t)$, the ion density in the region of the cloud to be considered. Unfortunately this cannot be solved rigorously since $n(t)$ is determined by the recombination equation:

$$dn/dt=\gamma-\alpha n^2-\beta n\int_0^\infty q^2N_qdq, \quad (10)$$

assuming, for the time being, equal numbers of positive and negative ions present. Again γ is the rate of formation of ion pairs per cubic centimeter and α the ordinary coefficient of recombination. α is very small so that in the absence of charged dust, n becomes fairly large until equilibrium is reached, but if charged particles are suspended in the air, the value of n is correspondingly depressed. Only certain extreme cases can be treated in more detail. If the aerosol is very dilute, the last term in Eq. (10) is very small compared to αn^2, and n will be practically the same as in dust-free air. It can there-

fore be regarded as constant in time if equilibrium between a constant ion production and recombination is established. This case is very simple since the integral in Eq. (9) is trivial.

On the other hand, if q is large and N_q not very small, the last term in Eq. (10) will easily predominate so that αn^2 effectively can be neglected. This means that practically all ions formed will attach to the smoke particles. It also presupposes that the suspended particles carry both positive and negative charges in about equal amounts, a case which seems to be frequently realized.[3] Moreover, if we make use of the fact that in this simplified case the rate of change of ion concentration is small compared to the ion production, we can reduce (10) to

$$0 \cong \gamma - \beta n \int_0^\infty q^2 N_q dq, \tag{11}$$

whence

$$\varphi = \frac{\gamma q^2}{\int_0^\infty q^2 N_q dq} = -\frac{dq}{dt}. \tag{12}$$

The integral in Eq. (12) is a function of time and since the integrand is in general not known in detail, further

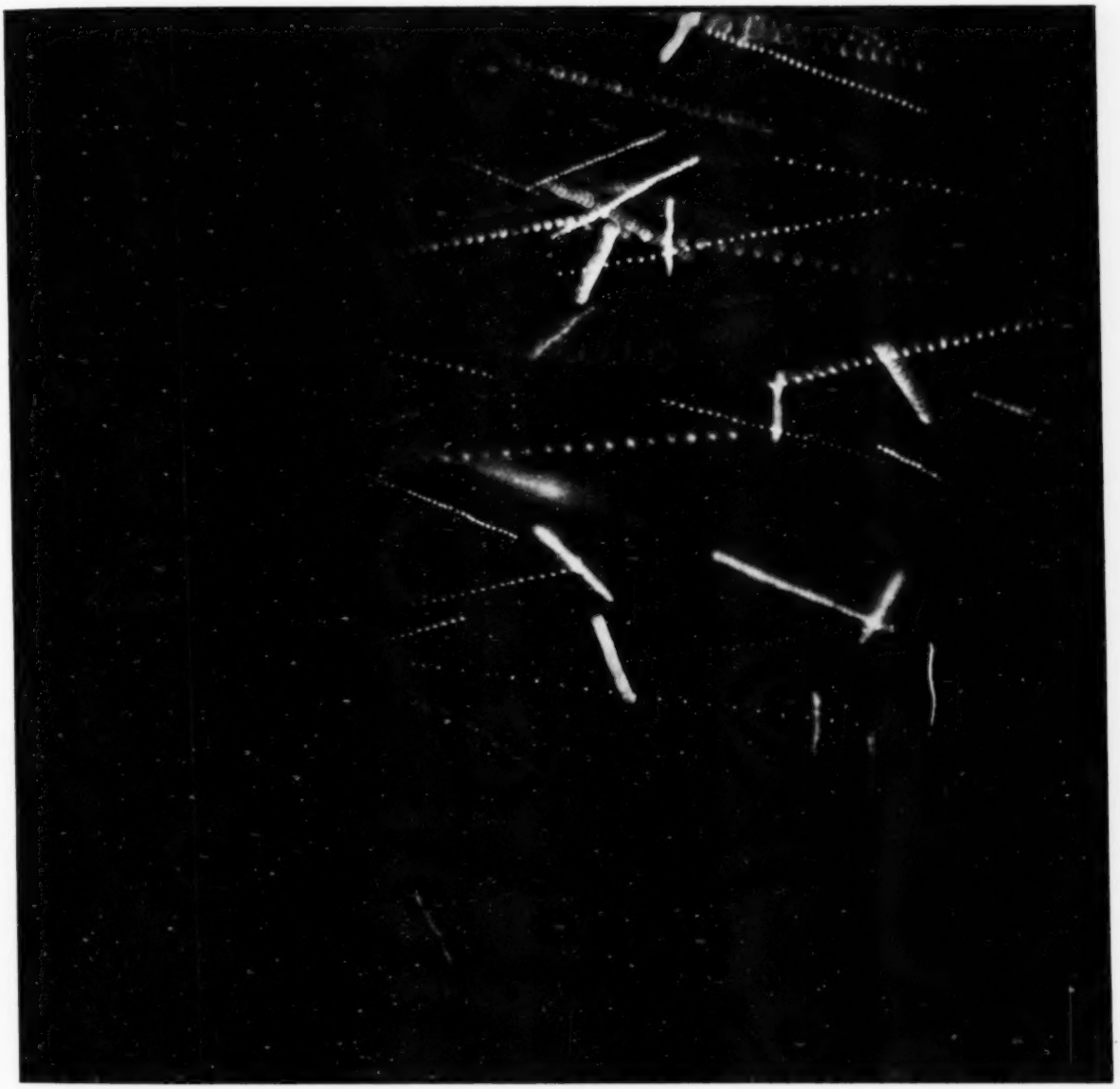


FIG. 1. Traces of quartz dust, 20 hours old, settling in air in a horizontal field of 560 volts cm^{-1}. Dark field illumination, 450 flashed min.$^{-1}$. Total exposure about 4 sec. Neutral particles settle vertically. Singly and doubly charged particles of both signs are easily recognized. For higher multiplicities the traces have to be carefully measured, for possible convection correction has to be made. Note the distinct effect of Brownian motion or differential heating on these irregularly shaped particles.

generalized development is impossible. However, by defining a root mean square charge

$$q^*(t)=\left[1/N\int_0^\infty q^2N_q dq\right]^{\frac{1}{2}} \tag{13}$$

and substituting this into (12), inspection shows that it is justifiable to write for the characteristic charge of the cloud $(dq^*/dt)=(-\gamma q^*/Nq^{*2})=-\gamma/N$ which yields

$$q^*=q_0{}^*-\gamma/Nt. \tag{14}$$

In other words, the r.m.s. charge in a symmetrically charged smoke cloud decreases linearly with time as long as the approximations introduced are justified. This is a rather remarkable result since the average charge of the smoke particles,

$$\bar{q}=(1/N)\int_0^\infty qN_q dq=Q/N \tag{15}$$

by our own definition decreases at the same rate:

$$d\bar{q}/dt=1/N(dQ/dt)=-\gamma/N,$$

where Q is the total charge of one sign per unit volume of the cloud.

Combining Eqs. (12)–(14), the most probable rate of change of charge of each particle is expressed as a function of its charge q and of the time, containing the initial r.m.s. charge $q_0{}^*$ of the cloud element as a parameter:

$$\frac{dq}{dt}=\frac{-\gamma q^2}{N[q_0{}^*-(\gamma/N)t]^2}. \tag{16}$$

This is readily integrated, resulting in an expression analogous to (7), to wit:

$$q=\frac{q_0}{1+(\gamma q_0 t/Nq_0{}^{*2}-\gamma q_0{}^* t)}, \tag{17}$$

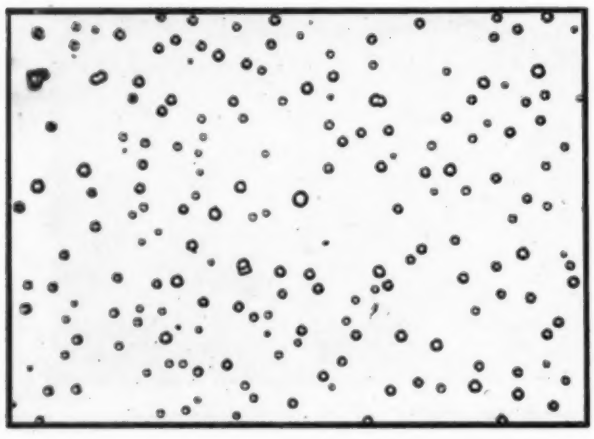


FIG. 2. Ammonium chloride smoke particles collected on microscopic cover glass at bottom of the dusting chamber. Sizes range from 0.7 to 3 microns in diameter.

or

$$q_0=\frac{q}{1-(\gamma qt/Nq_0{}^{*2}-\gamma q_0 t)}, \tag{18}$$

which has to be substituted in Eq. (8).

Unfortunately we have not been able to verify these simple relations quantitatively since, in our experimental arrangement,[5] it was not possible to keep a smoke homogeneously suspended for a sufficiently long period. Qualitatively, however, these ideas appear to be in good agreement with observations discussed elsewhere.[6]

EQUILIBRIUM DISTRIBUTION

The approximations introduced above are justified only if the charges are fairly large, say, with an average of 50 or more electrons per particle, and if the aerosol is not too dense. As the charges of the particles decrease, the ion-ion recombination gains in importance so that Eq. (11) can no longer be used and the second treatment breaks down. All we can say is that although the ion concentration will gradually increase, the rate of neutralization of the smoke will decrease more and more until φ_2 becomes comparable to φ_1 so that the latter no longer can be neglected. In other words we have to return to Eq. (1). Moreover we shall no longer be able to make use of differentials so that a rigorous treatment would be extremely complicated. Again, only the very extreme case of negligible change of charge distribution can be sufficiently simplified to permit a few quantitative conclusions. In this case the conditions of balance tell us that not only $\Delta N_q=0$ but that the terms on the right hand side of Eq. (1) have to cancel in pairs, i.e.,

$$N_{q-1}\varphi_1(q-1)-N_q\varphi_2(q)=0, \tag{19}$$

if the equilibrium is to be stable. Thus the equilibrium distribution can be expressed in the form

$$N_q=N_0\prod_{l=0}^{l=q}\frac{\varphi_1(l-1)}{\varphi_2(l)}, \tag{20}$$

if we define $\varphi_1(-1)=\varphi_2(0)=1$. Since φ_1 and φ_2 depend on the size of the particles, each size range will have a different distribution unless the size dependence cancels out in the expression. It is clear that $\varphi_1(q)$, the probability of charge addition, will decrease as q increases. The question arises whether $\varphi_1(2)$ for most smokes is not already negligible so that practically no particles with more than two elementary charges will be present in the cloud near equilibrium. Unfortunately many factors contributing to φ_1, such as the effect of the material and the shape, are largely unknown so that quantitative estimates cannot be expected to be reliable. We can assume, however, that the charges on a spherical

[5] W. B. Kunkel and J. W. Hansen, Rev. Sci. Inst. **21**, 308 (1950).

[6] W. B. Kunkel, J. App. Phys. **21**, 820 (1950).

particle arrange themselves on its surface at maximum equal distances (conducting droplets) and that all ions actually colliding with the surface will impart their charge to it. This very crude first approximation leads to the conclusion that in a smoke consisting of particles of about one micron diameter, multiple charges as high as five or even ten electrons are still quite likely.

This rather unexpected result is in qualitative agreement with the following experimental evidence. A small amount of fine quartz powder was dispersed by a short blast of air. The particles of the dust cloud thus generated were found to be quite highly charged, positive and negative charges being present in about equal quantity. The charges and sizes of the particles were determined by photographing their trajectories as they settled in a horizontal electric field with dark field illumination. The time scale was provided by using intermittent light of known flash frequency. This method has been described in some detail elsewhere.[5] After about one hour, all particles larger than one micron in diameter had settled out. Most of the small particles still present carried charges of between 10 and 20 electron units of either sign, the total range being from zero to about 100 electron units. A minute convection kept most of these particles suspended for 24 hours, after which time the charges were measured again. It was found that the average charge had decreased to about 3 electron units, charges of higher multiplicity still being present in large number. An example of the pictures of the falling particles obtained is shown in Fig. 1. There appeared to be no significant change in this distribution during the following few hours for which the suspension could be maintained.

In order to confirm the theoretical considerations above, it was next demonstrated that the charges observed were not merely residues of the initial contact electricity but represented a true equilibrium condition. A cloud of ammonium chloride smoke was generated by mixing ammonia and hydrochloric acid vapors in room air of 40 percent relative humidity at 20°C. The particles formed under suitable conditions were fairly uniform in size and quite round as seen in Fig. 2. Apparently aggregation between large particles did not take place, the growth consisting entirely of addition of very small particles or of single gas molecules of NH_4Cl to solutions of NH_4Cl in water. The surfaces appeared to be quite smooth and spherical under the electron microscope. A few minutes after complete formation of the cloud, analysis showed these particles to be mostly neutral, a certain fraction being singly charged. Three hours later the percentage of charged particles had markedly increased, multiple charges being already present in appreciable number. Again a part of the smoke was kept suspended by a slight convection in the chamber. The distribution found on the next day is shown in Fig. 3. The particles were grouped according to their charge irrespective of their size. Actually the average charge per particle increased somewhat with

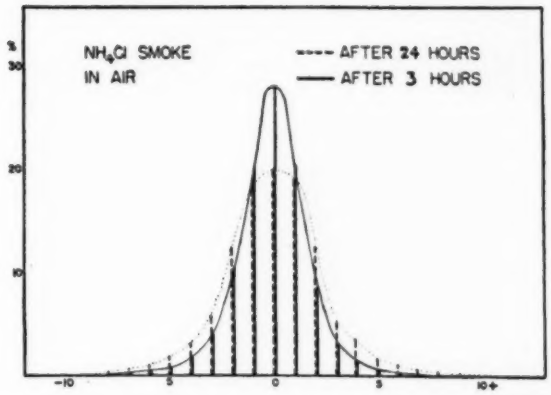


FIG. 3. Charge distribution in ammonium chloride smoke made by mixing the gaseous components. The curves are based on totals of about 1000 particles each.

the size of the particle, but since the numbers were not large enough to provide reliable statistics for separate size-groups, the all-over distribution for all particles together is shown. The general picture is very similar to the case of the quartz dust which started out charged, indicating that some kind of an equilibrium can actually be reached and is arrived at either by neutralization or by charging. It is seen that the number of singly charged particles is very little less than that of the neutral ones in agreement with the argument that $\varphi_2(1)$ can be only very little larger than $\varphi_1(0)$ for particles of this size. However, the observed curve falls off much faster than would be predicted by the simple theory indicated above. The most obvious explanation of this deviation is that ions colliding with particles do not necessarily impart their charge to the latter, especially not if the particle is already carrying an excess of the same sign as the ion. Nevertheless it is quite certain that an appreciable percentage of smoke particles in equilibrium are carrying multiple charges. The largest charge observed by us on an ammonium chloride particle of one micron diameter one day after formation was 18 electrons. It is interesting to note that this direct verification of the existence of multiply charged smoke particles in equilibrium with ionized air is a rather nice verification of the findings of G. Vassails,[7] who ascribed the appearance of groups of Langevin ions of somewhat smaller size to the multiple charges acquired from the natural ions in the air. The fact that the distributions always were practically symmetrical seems to indicate that both photoelectric effect and direct ejection of electrons by corpuscular radiation are negligible under normal conditions.

The work on dust electrification of which this is an interesting by-product was undertaken on an ONR contract at the suggestion and under the supervision of Professor L. B. Loeb. The writer desires to acknowledge his appreciation of the advice and encouragement given him in various phases of this work by Professor Loeb.

[7] G. Vassails, thesis, Masson, Paris (1948).

Double Crystal and Slit Methods in Small Angle X-Ray Scattering*

H. N. Ritland,† P. Kaesberg, and W. W. Beeman
Department of Physics, University of Wisconsin, Madison, Wisconsin
(Received February 17, 1950)

An angular resolution of about 10 seconds of arc may be obtained with a double calcite crystal spectrometer for Cu $K\alpha$-radiation. However, the shape of the rocking curve at large angles is such as to make the detection of radiation from weakly scattering materials difficult. Of the order of 10^6 counts per second are detected with a stationary, water-cooled copper target and parallel crystals. Wing intensities, even at large angles, do not fall much below 10^2 counts per second. Wing heights may be considerably reduced by the addition of a third calcite crystal. With this modification the crystal spectrometer is a useful instrument for most scattering experiments in which extremely high resolution is desired. It is inferior to a slit system and Geiger counter arrangement when poorer resolution may be tolerated but high sensitivity is required. A slit instrument with an angular resolution of 4 or 5 minutes of arc provides an effective incident flux of more than 10^6 counts per second while wing intensities may be as low as 1 count per second.

INTRODUCTION

THE study of small angle x-ray scattering involves a number of experimental difficulties in much more severe form than in conventional x-ray diffraction work. Among these are the requirements of high resolution and high sensitivity as well as good monochromatization and accurate measurement of intensity. These difficulties arise from the necessity of obtaining quantitative information about the rapidly changing scattered intensity in a narrow angular region about the primary beam.

In most small angle scattering studies higher resolution and sensitivity are obtained by making suitable refinements in photographic x-ray diffraction technique, principally involving finer slits and larger specimen-to-film spacing.[1] The diffraction chamber is usually evacuated to eliminate air scattering. Monochromatization is obtained by using radiation reflected from single crystals or by using selective filters. Accurate intensity measurements can be made only with difficulty because of the limited range of linear response of photographic film.

This paper discusses and compares the properties of two instruments which have been found useful for the study of small angle scattering, both involving Geiger counter detection. One utilizes crystals for collimation, while the other uses slits. The crystal apparatus has extremely high resolution, but has comparatively low effective sensitivity due to a background caused by diffuse scattering from the crystal surfaces. The slit apparatus has less resolution, but has high sensitivity.

The relative importance of resolution and sensitivity depends upon the scattering material being studied. A recently completed experimental study[2] of several proteins in solution illustrates the case where resolution requirements are not excessive but high sensitivity is required. Dexter and Beeman[3] describe a method for obtaining particle sizes from multiple scattering. Their method requires the resolution provided by the crystal apparatus.

TWO AND THREE CRYSTAL SPECTROMETERS

The two crystal spectrometer[4] (as well as the three crystal instrument to be described later) is shown schematically in Fig. 1. The first calcite crystal collimates the beam and is fixed in position so that line radiation from the x-ray tube strikes it at the Bragg angle. The scattering material is placed midway between the two crystals. The angular distribution of the radiation leaving the scatterer may be analyzed by rotating the second calcite crystal, since its reflectivity is high only for radiation impinging upon it at the Bragg angle. As the second crystal is rotated, that part of the scattered beam which will be reflected, does not continue to strike the same point in the crystal. As a result, the scattered radiation leaving the second crystal is focused to a point beyond it, at a distance equal to that between the second crystal and the scatterer. A thin window argon-alcohol Geiger counter is placed at this point and remains stationary as the second crystal rotates.

Slits placed at the tube window and before the sample sharply delimit the wave-length spread of the continu-

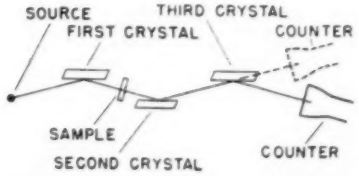


FIG. 1. The geometry of the crystal spectrometers for small angle scattering. In the double crystal arrangement only the first two crystals are used, and the counter position is shown by the dashed line drawing. The drawing has been distorted slightly in that the counter has been placed somewhat too far away from the second crystal. The full line drawing of the counter indicates its position for the three crystal arrangement.

* This work was supported in part by the joint program of the ONR and AEC and in part by the Wisconsin Alumni Research Foundation.

† Now at the University of Washington, Seattle, Washington.

[1] O. E. A. Bolduan and R. S. Bear, J. App. Phys. **20**, 983 (1949).

[2] Ritland, Kaesberg, and Beeman, "An x-ray investigation of the shapes and hydrations of several protein molecules in solution" (to be published).

[3] D. L. Dexter and W. W. Beeman, Phys. Rev. **76**, 1782 (1949).

[4] Kaesberg, Ritland, and Beeman, Phys. Rev. **74**, 71 (1949).

ous spectrum striking the scatterer so that its effect is negligible compared with the desired line radiation. The $K\beta$-radiation may be eliminated by the proper filters. Second-order radiation is passed by the crystal and must be guarded against when strongly absorbing samples are used. However, in ordinary cases where the absorption of the line radiation is about a factor of three, second-order intensity is entirely negligible since it is only of the order of 1/1000 of the intensity of the $K\alpha$-lines at 30 kilovolts, and may be reduced further by decreasing the voltage. In addition, the Geiger counter, when filled to an argon pressure sufficient to absorb 90 percent of the first-order radiation, absorbs but 25 percent of second order.

With perfect calcite crystals and copper radiation the experimental half-width at half-maximum for a double crystal rocking curve (intensity *vs.* angle) is known to be about 5 seconds of arc. This is in good agreement with theoretical calculations.[5] Experimental rocking curves are available in the literature only for the central portion of the curve, but theory predicts that they decrease inversely as the angle squared for large angles.[6] However, it is not known to what angles this type of behavior persists in real crystals. In order to evaluate the advantage of the narrow width of the double crystal rocking curve, detailed information concerning the wing intensity at large angles must be had. For diffuse scattering the area of the central part of the rocking curve should be compared with the wing intensity, while for line scattering the significant ratio is the height of the rocking curve at the center to the wing height at the position of the scattered line.

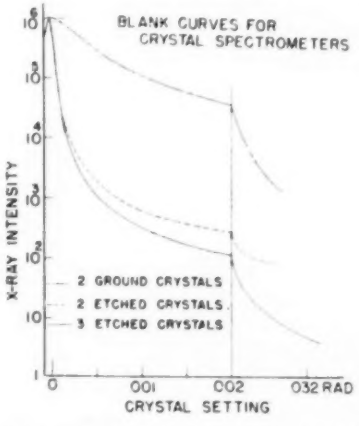


FIG. 2. The top curve is the double crystal rocking curve for a pair of calcite crystals that have been ground with carborundum to increase their reflectivity. The second curve shows the rocking curve obtained when these crystals were thoroughly etched with 0.5N hydrochloric acid. The lower curve shows the three crystal rocking curve that results when a third etched crystal is added (as described in the text). To facilitate comparison of their shapes, the three curves have been normalized to the same ordinate at zero rocking angle.

[5] A. H. Compton and S. K. Allison, *X-Rays in Theory and Experiment* (D. Van Nostrand Company, Inc., New York), second edition, pp. 709–730.

[6] F. P. Miller, Phys. Rev. **56**, 757 (1939).

TABLE I. The table shows the effect of etching treatment with 0.5 N HCl on the shape of the rocking curve for a pair of ground calcite crystals. The intensities and the areas are in arbitrary units. Copper radiation was used.

	A Etching time	*B* Center intensity	*C* Rocking curve width	*D* Rocking curve area[a]	*E* Wing height at 1°	*F* Ratio *E/B*	*G* Ratio *E/D*
1.	0 sec.	100,000	116 sec. arc	130	140	$140 \cdot 10^{-4}$	1.1
2.	5	200,000	96	220	700	350	3.2
3.	10	250,000	60	175	300	120	1.7
4.	15	240,000	30	95	90	38	1.0
5.	20	230,000	16	45	45	20	1.0
6.	25	230,000	14.5	40	20	9	0.5

[a] Within 200 seconds of arc of zero rocking angle.

We have found that curves obtained with most calcite crystals as cleaved, or ground and sufficiently etched, very nearly equal the theoretical width and indeed have wings which decrease in intensity inversely as the angle squared to about 3 or 4 minutes of arc. But the wings then go down more slowly and vary considerably with the choice of crystals and with treatment of their faces. To determine some of the effects of crystal surface conditions upon the usefulness of the instrument, a series of rocking curves was taken with a pair of ground crystals (ground with No. 600 carborundum) as a function of the time of etching in 0.5 normal hydrochloric acid. Table I summarizes the results and Fig. 2 shows rocking curves for these crystals before and after thorough etching.

It may be seen that with a light etch the central intensity is considerably increased but then remains substantially constant with further etching. The wing heights are increased at first, and then reduced as the etching proceeds, but the reduction has become appreciable only after the central area of the rocking curve is small compared with its original value. It is thus apparent that the reflected intensity may be increased by misaligning some of the surface layers, but only at the expense of a great loss in sensitivity because of the increased wing background. These results strongly suggest that best experimental conditions occur when the crystals have been etched enough to obtain as narrow a rocking curve as possible.

If a third crystal is placed so that it is always parallel to the second crystal (and the two rotated together) it should reflect almost no radiation that leaves the second crystal at other than the Bragg angle, but should reflect Bragg radiation with little diminution of intensity. Such a modification is shown schematically in Fig. 1. With this modification one no longer has a focal point at which to place the Geiger counter, so that a counter with a sufficiently large sensitive volume must be used, or else the counter must be moved as well as the second and third crystals.

The addition of a third crystal does not greatly reduce the central (i.e., all three crystals parallel) intensity of the rocking curve since the radiation has already been highly collimated by the first two crystals. We observed reductions of central intensity by a factor of between two and three upon adding a third crystal. This is a

TABLE II. A comparison of the shape of the rocking curves for two and three crystal arrangements. The curves are normalized to 1,000,000 at the center. Iron radiation was used. These curves are for well-etched, near perfect crystals.

Seconds of arc from the parallel position	Intensity 3 crystals	Intensity 2 crystals
0	1,000,000	1,000,000
25	23,000	35,000
50	5,300	8,700
100	1,000	2,300
400	100	280
1,000	33	150
4,000	8	110
8,000	3	100

slightly smaller percent reflection than theory would predict for perfect calcite crystals.

However one expects the wings to be reduced about twice as much as the central intensity with the addition of a third crystal parallel to the second. This results from the fact that when the first two crystals are not parallel the radiation leaving the face of the second can be divided into two equal parts one of which is strongly and the other very weakly reflected by the third crystal. The two parts are a component reflected at the center of the single crystal diffraction pattern by the first crystal but at the wing by second crystal and a component reflected on the wing by the first crystal but the center by the second. Only the latter component is appreciably reflected by the third crystal. It is diminished by the third reflection by about the same factor as is the central intensity when all three crystals are parallel.

The three crystal rocking curve also should have wings diminishing as the reciprocal of the square of the angle from parallelism.

A glance at Table II shows that the three crystal wings go down more slowly than the predicted $1/\varphi^2$ at large angles but that the improvement on the two crystal arrangement may still be much greater than a factor of two. At the largest angles the three crystal wings are less than the two crystal by more than an order of magnitude. This implies that no appreciable fraction of the wing intensity above the theoretical $1/\varphi^2$ contribution leaves the second crystal at the Bragg angle.

The cause of the excess wing intensity is not clear. One of the components leaving the second crystal is an intense beam from the first crystal which is presumably only weakly reflected by the second. This beam may be the source of temperature diffuse scattering at the second crystal or of a broad scattering peak from very small crystalline domains.

Although the third crystal has a pronounced effect in reducing wing heights, they are still about 10 times room background for large wing angles (with a stationary copper target tube). Thus the sensitivity of the three crystal spectrometer, while higher than the two crystal instrument, is still inadequate for diffuse scattering studies with materials of low scattering power. Since the scattered intensity depends upon the square of the difference in electron density between the particles and their surrounding medium, this lack of sensitivity will exclude the important class of experiments in which the scattering specimens are biological materials in aqueous solution.

SLIT COLLIMATION APPARATUS

If slits, rather than crystals, are used for collimation, resolution, as well as intensity, is governed by the slit widths and separations. As a result, for a given distance between two collimating slits, the highest attainable resolution is determined by the lowest tolerable intensity. Thus, the Geiger counter, while not an essential feature of a slit collimating system (as it is in the crystal spectrometers), may be used to advantage because it has high efficiency for detecting radiation.

The instrument described in the succeeding paragraphs was designed to study the angular distribution of scattered x-rays from particles in the size range 20A to 200A in diameter. The slit geometry was such that a convenient level of scattered intensity could be maintained with weakly scattering protein solutions as scattering specimens and with a water-cooled, stationary copper or iron target, x-ray tube as the radiation source.

The geometry of the slit system is indicated schematically in Fig. 3. A pair of stationary tantalum slits collimates the beam incident upon the scatterer. A second pair of slits analyzes the angular distribution of the radiation leaving the scatterer. The second pair of

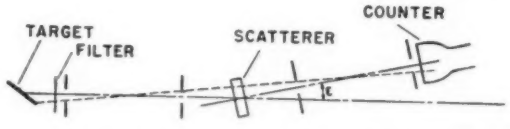


FIG. 3. A schematic drawing of the slit system described in the text. The scale has been greatly compressed in the horizontal direction. The angle ϵ is the smallest rocking angle at which the counter no longer sees the target.

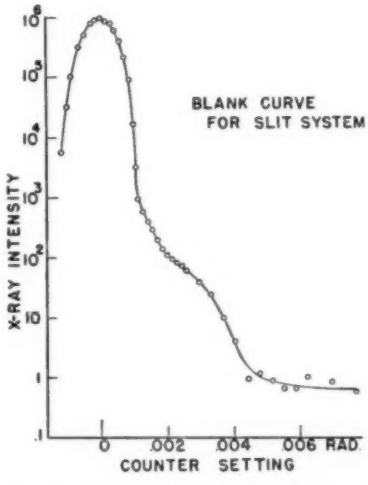


FIG. 4. The blank rocking curve for the slit system described in the text. The curve shown is for slit widths of 0.02 cm. The $4d/a$ angle is then 0.0027 radian.

slits, and also the specimen holder and the Geiger counter, are fixed to an arm which may be rotated, by means of a calibrated screw, about an axis through the center of the specimen. The distance between successive slits is 30 cm and the scatterer is halfway between the second and third slits. The slits are 1 cm high, but can be cut down by suitable apertures. The slit widths are adjustable, usually being between 0.1 and 0.4 mm wide. In order to reduce the loss by air absorption, the slits are mounted in two evacuated tubes with windows of 0.001-in. Nylon film at each end. Monochromatization is achieved by means of the Ross filter method. The differential filters are mounted on a slide in front of the x-ray tube exit window. The scatterer, Geiger counter, and filters are not in the evacuated region.

If intensity *vs.* angle is plotted for the slit apparatus with no scattering sample, a curve is obtained that is analogous to the rocking curve for two and three crystal spectrometers. This curve is shown in Fig. 4. The central part of the curve consists mainly of radiation that goes straight through all of the slits. Wing radiation arises from slit edge scattering, and from air scattering in the non-evacuated regions. If the slit width is d and the separation is a, the counter no longer sees the x-ray tube target at an angle of rotation of $2d/a$ from the straight through position (as shown in Fig. 3). Scattering from the edges of the second and third slits is eliminated at an angle of $4d/a$. Air scattering from the non-evacuated region around the sample falls off rapidly with angle, because the region where the two telescopes overlap decreases rapidly.

The most striking feature of the slit rocking curve (Fig. 4) when compared with the crystal rocking curves (Fig. 2) is its exceedingly low wing intensity. At angles only slightly greater than $4d/a$, the ratio of center to wing intensity is greater than 10^6. The three crystal rocking curve approaches this ratio only at comparatively large angles (see Fig. 2) while the two crystal curves are two or three orders of magnitude worse.

Under normal operating conditions a slit instrument with an angular resolution of 4 or 5 minutes of arc provides an effective incident flux of about 10^6 counts per second. This is approximately the counting rate for the three crystal spectrometer at zero rocking angle. However, a more significant comparison for diffuse scattering is the area under the central portion of the rocking curve for each instrument. The slit system curve area exceeds that of the three crystal instrument by a factor of 25.

We wish to thank the Kimberly-Clark Corporation for their support of the Kimberly-Clark Fellowship held by one of us (H.N.R.) during the year 1948–49.

Letters to the Editor

Vacuum Tubes with Mutually Bombarding Oxide Cathodes

E. G. Hopkins

Electrical Engineering Department, The New South Wales University of Technology, Sydney, New South Wales, Australia

March 13, 1950

EXPERIMENTS in which two flat oxide cathode surfaces 3.5 cm long and 1.0 cm wide were mounted face to face to form a high vacuum diode and were caused to act alternately as anode and cathode in passing a current of 50 cycles per second between them have shown that such an arrangement can be made to operate stably after the removal of the initial heating source. If a 240-volt a.c. supply is connected to the cathodes through a series resistance at an instant when the temperature of their hottest parts is above a critical minimum (below the visible range), the current builds up to a value determined by the space-charge limited characteristics of the diode. One diode with a spacing of 3 mm was found to obey closely the law, r.m.s. amp. $=1.9$ (r.m.s. volts)$^{1.42}\times10^{-4}$ between 120 and 220 volts, the corresponding range of maximum cathode brightness temperatures measured on an uncoated nickel spot being 880 to 1260°K. The current wave form was quite symmetrical about the zero axis and showed no signs of saturation. Portions of the cathode below a visible temperature glowed with a pale blue fluorescence under bombardment. One sealed-off diode was allowed to run for seven hundred hours.

A triode was constructed consisting of two cathodes as described having a total separation of approximately 2 mm and a grid of molybdenum wires spaced equally between them. Tests were made with a grid voltage applied through a synchronous switch in such a way that for any half-cycle it consisted of a half-sine wave applied negatively relative to the particular electrode then acting as a cathode. By assuming the Langmuir equation for triode current to hold for all instantaneous voltages, the equation I_p (r.m.s.) $= K$ (E_b r.m.s. $+ \mu E_g$ r.m.s.)^n was used to estimate μ, the amplification factor from values of voltage and current read on r.m.s. reading instruments. This equation is useful when examining the possible use of such a triode to control a.c. power circuits.

The triode was made to oscillate at a frequency of 0.5 Mc/sec. in an oscillator circuit which included a synchronous switch to reverse the phase of the grid drive. When operating stably with a cathode brightness temperature of about 950°K, the oscillator drew 0.12 amp. at 220 volts and delivered approximately 5 watts to the load lamp.

These experiments demonstrate the possibility of utilizing the power usually wasted in anode dissipation to heat a cathode surface much larger than can be provided economically in conventional power tubes. It appears feasible from the data obtained from this work and from experience with conventional tubes that if the spacings of a practical tube based on the above principle were made as small as possible without causing excessive grid emission, hot spot formation, or mechanical troubles, then such a tube could be made to deliver useful amounts of high frequency power at acceptable efficiency without the use of transformers or rectifiers when operated directly from an a.c. supply of 240 volts or over.

These experiments, details of which will be published when the present work has reached a more advanced stage, were carried out by the writer at the Radiophysics Vacuum Laboratory in the Electrical Engineering School of the University of Sydney.

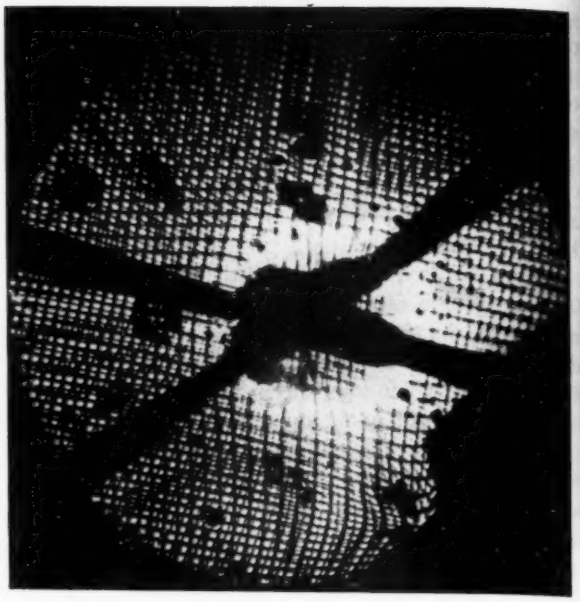


FIG. 1.

Electron Optical Properties of Space-Charge Clouds

L. MARTON AND D. L. REVERDIN
National Bureau of Standards, Washington, D. C.
April 13, 1950

THE space-charge lens has been repeatedly considered as a corrective element for the reduction of the aperture defect of electron lenses.[1] This is due to the recognition by Scherzer[2] that no axially symmetrical, time invariant, space-charge free field is free from the aperture defect, and that dropping of any of these conditions may improve the behavior of the lens. Gabor,[3] who was one of the first to propose the use of space-charge clouds as corrective elements, raised, however, a question and suggested an experiment: "Will the 'space-charge' cloud not act on the beam like frosted glass?" He proposed "to settle this point by the simple experiment of shooting a fast electron beam parallel to the axis through a static magnetron. It may be mentioned that such an experiment would also at once settle the question how nearly Hull's simple theory agrees with actual conditions in a magnetron as, on Hull's theory, the beam would suffer no radial deflection."

Gabor's proposed experiment forms the subject of a dissertation for a Ph.D. degree by one of us[4] and will be published separately. The present brief communication deals with a partial answer to Gabor's question about the suitability of the space-charge cloud as a lens-correcting element.

In the present experiment, a wide electron beam was sent axially through a static magnetron. Before reaching the magnetron, the beam passed one wire mesh, and a second wire mesh was placed in the beam's passage after the magnetron. The details of the experimental conditions are given in reference 4. A typical observation is shown in Fig. 1. This figure represents the photographed shadow patterns cast on a fluorescent screen. The shadows of the two wire meshes are plainly visible and in addition to them, there are two heavier shadows shown at approximately a 60° angle. These latter are the shadows of the filament supporting bars. Very marked in Fig. 1, also, is the bright ring located between the cathode (in the center of the picture) and the anode (dark circular shadow at the circumference of the picture). The outer edge of this bright ring lies approximately where the theoretical electron cloud radius would be according to the classical magnetron theory.

Approximate computations carried out on this pattern (Fig. 1), show that the bright ring is produced by a high intensity space-charge cloud acting as a divergent lens. Although experimental conditions prevented us from actually measuring the focal distance of this divergent lens, this latter could be estimated to be rather large. A comparison of the pattern shown in Fig. 1, with patterns taken in absence of the space charge, shows that a partial correction of the aperture defect is probably present, although we do not have quite definite indications due to the preliminary nature of this investigation.

The wire shadows in the bright region of the picture appear much less sharp than toward the periphery. When placing an oscilloscope in the plate circuit of the magnetron, we could observe that whenever the electron optical observation showed the presence of space charge, considerable random noise could be observed in the plate circuit. Although our experimental set-up does not exclude entirely the possibility of very high frequency oscillations, the probability of such oscillations being present is extremely reduced. Thus the lack of definition in the bright region can be fairly well correlated with the noise and the conclusion drawn that the image definition is reduced by random noise in the space charge.

The present experiments thus seem to indicate that, whereas a correction of the aperture defect may be possible by means of space charge, Gabor's suspicion of its acting as a "frosted glass" seems to be justified. The present observations, being mostly an outgrowth of the magnetron investigation, are not presented as arguments for abandoning all attempts toward correction of aperture defect by means of a space-charge cloud. They seem to show, however, that the reduction of the resolution in the image may outweigh considerably whatever advantage a space charge could offer as a lens-correcting element.

[1] V. E. Cosslett, *Electron Optics* (Clarendon Press, Oxford, 1946), pp. 142–147.
[2] O. Scherzer, Optik **2**, 114 (1947).
[3] D. Gabor, *The Electron Microscope* (Chemical Publishing Company, Inc., New York, 1948), p. 124.
[4] Doctoral dissertation submitted by D. L. Reverdin in January, 1950, to George Washington University, Washington, D. C. (to be submitted in abbreviated version to J. App. Phys.).